

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number

WO 2017/184831 A1

(43) International Publication Date
26 October 2017 (26.10.2017)

(51) International Patent Classification:
C07K 16/00 (2006.01) *C12N 15/90* (2006.01)
C07K 16/46 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:
PCT/US2017/028552

(22) International Filing Date:
20 April 2017 (20.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/325,385 20 April 2016 (20.04.2016) US

(71) Applicant: REGENERON PHARMACEUTICALS, INC. [US/US]; 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US).

(72) Inventors: BABB, Robert; c/o Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US). BURAKOV, Darya; c/o Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US). CHEN, Gang; c/o Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US). FANDL, James, P.; c/o Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US).

(74) Agent: GROLZ, Edward, W.; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Suite 300, Garden City, NY 11530 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

WO 2017/184831 A1

(54) Title: COMPOSITIONS AND METHODS FOR MAKING ANTIBODIES BASED ON USE OF AN EXPRESSION-ENHANCING LOCUS

(57) Abstract: This invention relates to site-specific integration and expression of recombinant proteins in eukaryotic cells. In particular, the invention includes compositions and methods for improved expression of antibodies including bispecific antibodies in eukaryotic cells, particularly Chinese hamster (*Cricetulus griseus*) cell lines, by employing an expression-enhancing locus.

COMPOSITIONS AND METHODS FOR MAKING ANTIBODIES BASED ON USE OF AN EXPRESSION-ENHANCING LOCUS

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority from U.S. Provisional Application No. 62/325,385, filed April 20, 2016, the entire contents of which are incorporated herein by reference.

FIELD OF THE DISCLOSURE

[0002] This disclosure relates to site-specific integration and expression of recombinant proteins in eukaryotic cells. In particular, the disclosure relates to compositions and methods for improved expression of antigen-binding proteins such as bispecific antibodies in eukaryotic cells, particularly Chinese hamster (*Cricetulus griseus*) cell lines, by employing an expression-enhancing locus.

BACKGROUND ART

[0003] Cellular expression systems aim to provide a reliable and efficient source for the manufacture of a given protein, whether for research or therapeutic use. Recombinant protein expression in mammalian cells is a preferred method for manufacturing therapeutic proteins due to, for example, the ability of mammalian expression systems to appropriately post-translationally modify recombinant proteins.

[0004] Despite the availability of various expression systems, the challenge of efficient gene transfer and stability of the integrated gene for expression of a recombinant protein still exists. For long-term expression of a target transgene, one consideration is minimal disruption of cellular genes to avoid changes in the phenotype of the cell line.

[0005] Engineering stable cell lines to accommodate multiple genes for expression, such as multiple antibody chains as in multispecific antibodies, is particularly challenging. Wide variations in expression levels of integrated genes may occur. Integrating additional genes may lead to greater variation in expression and instability due to the local genetic

environment (*i.e.*, position effects). Expression systems for the production of multispecific antigen-binding proteins often requires the expression of two or more different immunoglobulin chains intended to pair as a specific multimeric format, and can often weigh in favor of homodimer production, rather than the desired heterodimer or multimer combination. Accordingly, there is a need in the art for improved mammalian expression systems.

SUMMARY OF THE DISCLOSURE

[0006] In one aspect, a cell is provided that contains an exogenous nucleic acid sequence integrated at a specific site within an enhanced expression locus, wherein the exogenous nucleic acid sequence encodes a bispecific antigen-binding protein.

[0007] In some embodiments, the exogenous nucleic acid sequence includes a first exogenous nucleic acid containing a nucleotide sequence encoding a first light chain fragment (LCF), a second exogenous nucleic acid containing a nucleotide sequence encoding a first heavy chain fragment (HCF), and a third exogenous nucleic acid containing a nucleotide sequence encoding a second HCF (or denoted as HCF* where the second HCF is different from the first HCF), wherein the first and second HCFs and the first LCF form a bispecific antigen-binding protein. In certain embodiment, the first and second HCFs and the first LCF contain at least two variable regions and two CH3 constant domains of a bispecific antigen-binding protein. In some embodiments, the two variable regions are different. In some embodiments, the two CH3 regions are different. In some embodiments, each exogenous nucleic acid sequence is integrated simultaneously at a specific site within the enhanced expression locus.

[0008] In some embodiments, the nucleotide sequence encoding the first HCF encodes amino acids from a first constant region (*e.g.*, encodes one or more of CH1, CH2, hinge or CH3 domain), and the nucleotide sequence encoding the second HCF encodes amino acids from a second constant region. The amino acids from a first constant region can be the same or different from the amino acids from a second constant region. In specific embodiments, the nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain, wherein the first and second CH3 domains can be the same or different. In some embodiments, the first and second CH3 domains differ in at least one amino acid position; *e.g.*, one of the two CH3

domains is a human IgG CH3 domain, and the other one is a modified human IgG CH3 domain, and the two CH3 domains have different Protein A binding characteristic. In other embodiments, the nucleotide sequences encoding the first and second CH3 domains differ from each other in that one of the nucleotide sequences has been codon modified.

[0009] In other specific embodiments, the nucleotide sequence encoding the first HCF encodes a first heavy chain variable (VH) region, and the nucleotide sequence encoding the second HCF encodes a second VH region, wherein the first and second heavy chains can have the same or different VH regions. In another embodiment, the first and second VHs can be linked to the same or different constant regions.

[0010] In some embodiments, the nucleotide sequence encoding the first LCF encodes a first light chain variable (VL) region.

[0011] In some embodiments, the exogenous nucleic acid sequence contains an additional exogenous nucleic acid that includes a nucleotide sequence encoding a second LCF, such as a second light chain variable (VL) region. In some embodiments, the nucleotide sequence encoding the second VL region also encodes a second light chain constant region.

[0012] The relative positions of the multiple exogenous nucleic acids at the locus can vary. In some embodiments, the LCF encoding nucleic acid is located upstream or downstream relative to both HCF-encoding nucleic acids.

[0013] In some embodiments, each of the HCF or LCF-encoding sequence is independently linked to a transcriptional regulatory sequence. In specific embodiments, the first exogenous nucleic acid further includes a first promoter operably linked to the nucleotide sequence encoding a first LCF, the second exogenous nucleic acid further comprises a second promoter operably linked to the nucleotide sequence encoding a first HCF, and a third exogenous nucleic acid comprising a third promoter operably linked to the nucleotide sequence encoding a second HCF, wherein the first, second and third promoters are the same or different, and/or said promoters are the same or different from the fourth promoter to which the fourth exogenous nucleic acid is operably linked. In some embodiments, the first, second and third promoters are the same.

[0014] In some embodiments, the exogenous nucleic acid sequence at the integration site further includes recombinase recognition sites, for example, a first recombinase recognition

site (RRS) located 5' relative to the first exogenous nucleic acid, and a second recombinase recognition site (RRS) located 3' relative to both the second and third exogenous nucleic acids, wherein the first and second RRSs are different. In some embodiments, a third RRS is also included and located 3' relative to the first exogenous nucleic acid, and 5' relative to one or both the second and third exogenous nucleic acids, wherein the third RRS is different from the first and second RRSs.

[0015] In some embodiments, the exogenous nucleic acid sequence can include a fourth exogenous nucleic acid containing a selectable marker gene. In specific embodiments, the fourth exogenous nucleic acid is located 3' relative to the first exogenous nucleic acid. In certain embodiments, the fourth exogenous nucleic acid is integrated as a split gene. In other embodiments, the fourth exogenous nucleic acid, or selectable marker, is located 3' of the third RRS, which is 3' of a fourth promoter which is operably linked to the fourth exogenous nucleic acid. In some embodiments, the selectable marker gene comprises the third RRS which has been inserted, optionally inserted within an intron of the selectable marker gene, wherein the third RRS is different from the first and second RRS.

[0016] In certain embodiments, the order of the exogenous nucleic acids at a locus can be: from 5' to 3', the first exogenous nucleic acid (encoding LCF), the fourth exogenous nucleic acid (encoding selectable marker), the second exogenous nucleic acid (encoding first HCF), and the third exogenous nucleic acid (encoding second HCF); and in some particular embodiments, the second exogenous nucleic acid contains a nucleotide sequence encoding a modified CH3 domain of a human IgG, and the third exogenous nucleic acid comprises a nucleotide sequence encoding the native CH3 domain of the human IgG.

[0017] In certain embodiments, the order of the exogenous nucleic acids at a locus is: from 5' to 3', the first exogenous nucleic acid (encoding LCF), the second exogenous nucleic acid (encoding first HCF), the fourth exogenous nucleic acid (encoding selectable marker), and the third exogenous nucleic acid (encoding second HCF), wherein the second exogenous nucleic acid comprises a nucleotide sequence encoding the native CH3 domain of a human IgG, and the third exogenous nucleic acid comprises a nucleotide sequence encoding a modified CH3 domain of the human IgG.

[0018] In some embodiments, the promoters linked to the HCF or LCF-coding sequences are the same, and are different from the promoter to which the selectable marker gene is operably linked.

[0019] In some embodiments, the bispecific antigen-binding protein specifically binds to a T-cell antigen and to a tumor cell antigen. Other suitable dual antigen specificities are also provided.

[0020] In some embodiments, the enhanced expression locus is selected from a locus comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1, or a locus comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 2.

[0021] In various embodiments, the cell is a CHO cell.

[0022] In another aspect, vectors are provided that are designed for site-specific integration of multiple exogenous nucleic acids.

[0023] In some embodiments, this disclosure provides a set of vectors which set includes a first vector containing from 5' to 3': a first RRS, a first nucleic acid containing a nucleotide sequence encoding a first LCF, and a third RRS; and a second vector containing from 5' to 3', the third RRS, a second nucleic acid containing a nucleotide sequence encoding a first VH region, a second RRS; wherein either the first or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF; and wherein the first and second HCF, and the first LCF, form a bispecific antigen binding protein.

[0024] In some embodiments, the nucleotide sequence encoding the second HCF is included in the first nucleic acid, optionally located downstream of the nucleotide sequence encoding the first LCF. In other embodiments, the nucleotide sequence encoding the second HCF is included in the second nucleic acid.

[0025] In some embodiments, the nucleotide sequence encoding the first HCF encodes a first chimeric constant region (e.g., encodes one or more of CH1, hinge CH2, or CH3 domain, or fragments thereof, from any isotype), and the nucleotide sequence encoding the second HCF encodes a second chimeric constant region. Examples of a chimeric constant region is described in PCT International Publication No. WO 2014/121087 A1, published on August 7, 2014, incorporated herein by reference. The amino acids from a first constant region can be the same or different from the second chimeric constant region. In specific embodiments, the

nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain, wherein the first and second CH3 domains can be the same or different. In some embodiments, the first and second CH3 domains differ in at least one amino acid position; e.g., one of the two CH3 domains is a human IgG CH3 domain, and the other one is a modified human IgG CH3 domain, and the two CH3 domains have different Protein A binding characteristic. In other embodiments, the nucleotide sequences encoding the first and second CH3 domains differ from each other in that one of the nucleotide sequences has been codon modified.

[0026] In other specific embodiments, the nucleotide sequence encoding the first VH region encodes a first heavy chain, and the nucleotide sequence encoding the second VH region encodes a second heavy chain, wherein the first and second heavy chains can have the same or different constant regions.

[0027] In some embodiments, the nucleotide sequence encoding the first LCF encodes a first light chain variable region (VL).

[0028] In some embodiments, each of the LCF- or HCF-encoding sequences is independently linked to a transcriptional regulatory sequence, such as a promoter. In specific embodiments, the promoter linked to a first HCF-coding sequence and the promoter linked to a second HCF-coding sequence are the same. In specific embodiments, the promoters linked to LCF(s) and HCF(s) are all the same.

[0029] In some embodiments, the first nucleic acid in the first vector further contains a 5' portion of a selectable marker gene, located at 5' to the third RRS in the first vector; and the second nucleic acid further contains the remaining 3' portion of the selectable marker gene, located 3' to the third RRS in the second vector - that is, the selectable marker gene is split into the two vectors. In other embodiments, selectable marker and the promoter to which it is operably linked are split between the two vectors, in other words the promoter and selectable marker gene are located on different vectors. In certain embodiments, the promoter operably linked to the marker gene is located in the first vector 5' to a third RRS, and the marker gene is located 3' of the third RRS in the second vector, and is 5' of a second promoter operably linked to a second nucleic acid and a third promoter operably linked to third nucleic acid. In some embodiments, the third RRS in the first vector is present within a 5' portion of an intron

of the selectable marker gene; and the third RRS in the second vector is present within a 3' portion of an intron of the selectable marker gene.

[0030] In specific embodiments, the first vector includes, from 5' to 3', the first RRS, the first nucleic acid, and the third RRS; and the second vector includes, from 5' to 3', the third RRS, the second nucleic acid wherein the second nucleic acid contains the nucleotide sequence encoding a first HCF and the nucleotide sequence encoding a second HCF, and a second RRS. In other specific embodiments, the first vector includes from 5' to 3', the first RRS, the first nucleic acid wherein the first nucleic acid comprises the nucleotide sequence encoding a first HCF and the nucleotide sequence encoding a second HCF region, and the third RRS; and the second vector includes 5' to 3', the third RRS, the second nucleic acid wherein the second nucleic acid comprises a nucleotide sequence encoding a first HCF region, and a second RRS. In any of these specific embodiments, the first nucleic acid can further include a 5' portion of a selectable marker gene, located at 5' to the third RRS in the first vector; and the second nucleic acid further comprises the remaining 3' portion of the selectable marker gene, located 3' to the third RRS in the second vector; and wherein optionally the third RRS in the first vector is present within a 5' portion of an intron of the selectable marker gene, and the third RRS in the second vector is present within a 3' portion of an intron of the selectable marker gene.

[0031] In some embodiments, the vector set can include an additional vector or vectors; for example, a vector containing one or more RRSs and a nucleotide sequence encoding a second LCF, or a vector encoding one or more recombinases that recognize the RRSs.

[0032] In other embodiments, this disclosure provides a vector designed to achieve site-specific integration of multiple exogenous nucleic acids via homologous recombination based on homology arms. In some embodiments, the vector contains an exogenous nucleic acid sequence that encodes a bispecific antigen-binding protein, flanked by a 5' homology arm and a 3' homology arm for integration into an expression enhancing locus of a cell.

[0033] In a further aspect, this disclosure provides a system that includes a combination of a cell and one or more vectors, and that can be utilized to make cells having integrated within an expression enhancing locus exogenous nucleic acids that together encode a bispecific antigen binding protein.

[0034] In certain embodiments, a system is provided that includes a cell and a set of vectors, wherein the cell contains, integrated within an enhanced expression locus of its genome a set of RRSs that are different from one another and spaced between one or more exogenous nucleic acids, such as selection markers, for recombinant exchange with genes of interest in a set of vectors; and wherein the RRSs in the set of vectors comprise the same arrangement as the RRSs in the cell.

[0035] In some embodiments, a system is provided that includes a cell and a set of vectors, wherein the cell contains, integrated within an enhanced expression locus of its genome from 5' to 3': a first RRS, a first exogenous nucleic acid, a second RRS, a second exogenous nucleic acid, and a third RRS, wherein the three RRSs are different from one another; wherein the set of vectors includes a first vector containing from 5' to 3', the first RRS, a first nucleic acid containing a nucleotide sequence encoding a first immunoglobulin chain or fragment thereof, and the second RRS; a second vector containing the second RRS, a second nucleic acid containing a nucleotide sequence encoding a second immunoglobulin chain or fragment thereof, and the third RRS; and wherein either the first nucleic acid or the second nucleic acid further includes a nucleotide sequence encoding a third immunoglobulin chain or fragment thereof. Upon introduction of the vectors into the cell, the first and second nucleic acids in the vectors integrate into the enhanced expression locus through recombination mediated by the first, second and third RRSs.

[0036] In some embodiments, the first exogenous nucleic acid in the cell contains a first selectable marker gene, and the second exogenous nucleic acid in the cell contains a second selectable marker gene, wherein the first and second selectable marker genes are different. The selectable markers exchange with the integrated exogenous nucleic acids in the cell.

[0037] In some embodiments, the first vector includes from 5' to 3', the first RRS, the first nucleic acid containing the nucleotide sequence encoding the first LCF, and the third RRS; and the second vector containing 5' to 3', the third RRS, the second nucleic acid, wherein the second nucleic acid containing both the nucleotide sequence encoding the first HCF and the nucleotide sequence encoding the second HC, and the second RRS. In other embodiments, the first vector contains from 5' to 3', the first RRS, the first nucleic acid containing the nucleotide sequence encoding the first LCF and the nucleotide sequence encoding the second

HCF, and the third RRS; and the second vector contains 5' to 3', the third RRS, the second nucleic acid containing the nucleotide sequence encoding the first HCF, and the second RRS.

[0038] In some embodiments, the first vector includes from 5' to 3', the first RRS, the first nucleic acid containing the nucleotide sequence encoding the first HCF, and the third RRS; and the second vector containing 5' to 3', the third RRS, the second nucleic acid, wherein the second nucleic acid containing both the nucleotide sequence encoding the first LCF and the nucleotide sequence encoding the second HCF, and the second RRS. In other embodiments, the first vector contains from 5' to 3', the first RRS, the first nucleic acid containing the nucleotide sequence encoding the first HCF and the nucleotide sequence encoding the second HCF, and the third RRS; and the second vector contains 5' to 3', the third RRS, the second nucleic acid containing the nucleotide sequence encoding the first LCF, and the second RRS. In any of these embodiments, the first nucleic acid in the first vector can further include a promoter located at 5' to the third RRS, and the second nucleic acid in the second vector further includes the selectable marker gene to which the promoter will be operably linked, located 3' to the third RRS. In other embodiments, the first nucleic acid in the first vector can further include a 5' portion of a selectable marker gene, located at 5' to the third RRS, and the second nucleic acid in the second vector further includes the remaining 3' portion of the selectable marker gene, located 3' to the third RRS; wherein optionally the third RRS in the first vector is present within a 5' portion of an intron of the selectable marker gene; and the third RRS in the second vector is present within a 3' portion of an intron of the selectable marker gene.

[0039] In some embodiments, the nucleotide sequence encoding the LCF is operably linked to a first promoter, the nucleotide sequence encoding the first HCF is operably linked to a second promoter, and the nucleotide sequence encoding the second HCF is operably linked to a third promoter, wherein the first, second, and third promoters are the same, and are different from the promoter to which a selectable marker gene, if present in one of the vectors, is operably linked.

[0040] In some embodiments, the nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain, wherein the first and second CH3 domains can be the same or different. In some

embodiments, one of the two CH3 domains is the native CH3 domain of a human IgG, and the other CH3 domain is a modified CH3 domain of the human IgG. In specific embodiments, the nucleotide sequence encoding the modified CH3 domain is in the first vector (i.e., the vector encoding the first LCF), optionally downstream of the nucleotide sequence encoding the first LCF. In other specific embodiments, the nucleotide sequence encoding the modified CH3 domain is in the second vector and is upstream of the nucleotide sequence encoding the unmodified CH3 domain.

[0041] In another aspect, this disclosure provides methods for making a bispecific antigen-binding protein.

[0042] In some embodiments, the method includes providing a system described herein that contains a cell having RRSs and a set of vectors containing multiple exogenous nucleic acids that together encode a bispecific antigen-binding protein and RRS matching the RRSs in the cell; introducing the vectors into the cell by transfection; selecting a transfected cell where the exogenous nucleic acids in the vectors have integrated into an enhanced expression locus of the cell through recombination mediated by the RRSs; expressing the polypeptides encoded by the nucleic acids in the transformed cell; and obtaining the bispecific antigen-binding protein from the transfected cell.

[0043] In some embodiments, the method can include a cell containing an exogenous nucleic acid sequence that encodes a bispecific-antigen binding protein integrated within an expression enhancing locus, expressing the bispecific antigen-binding protein from the exogenous nucleic acid sequence; and obtaining the bispecific antigen-binding protein from the cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0044] **Figure 1.** An exemplary bispecific cloning strategy for integration within an expression enhancing locus. A light chain ("LC") vector, *e.g.* a common light chain, and a dual heavy chain ("HC") vector (with "*" indicating that the two HCs are different, *e.g.* HC* contains a modification in the CH3 domain and/or is codon-modified) are made by cloning variable regions of antibody of interest into appropriate vectors. The 3' RRS site of the LC vector and the 5' RRS site of the dual HC vector are the same and included in a split intron of a hygromycin resistance gene, engineered to combine and excise the intron to allow

expression of the protein encoded by the hygromycin resistance gene for efficient selection of recombinants. The arrows represent promoters.

[0045] **Figure 2.** An exemplary bispecific cloning strategy for integration within an expression enhancing locus. Utilizing a universal light chain (see, *e.g.* from a Humanized Universal Light Chain (ULC) VelocImmune® mouse, as described in WO 2013022782) having a 5' RRS (RRS1) allows efficient construction of new bispecific antibodies by inserting one heavy chain (HC*) flanked by a third RRS (RRS3) into pre-existing plasmids containing expression cassettes for the universal light chain. The second heavy chain (HC) is cloned into a second plasmid with RRS2 and RRS3 sites.

[0046] **Figure 3.** An exemplary bispecific cloning strategy for integration within an expression enhancing locus. The three distinct antibody chains (AbC1, AbC2, and AbC3) of a bispecific antibody are first cloned into individual vectors. The AbC1 and AbC3 vectors each have RRS sites flanking the antibody expression cassette. Expression cassette for AbC2 is excised from the AbC2 plasmid and then subcloned into AbC3 expression plasmid, giving rise to a plasmid that contains, from 5' to 3', an RRS3 site, AbC2 expression cassette, AbC3 expression cassette, and an RRS2 site. This plasmid, together with the AbC1 plasmid and a recombinase, are introduced into host cells that harbor RRS1 and RRS2 in an expression enhancing locus. Bispecific antibody expression cell lines are isolated after recombinase-mediated cassette exchange.

[0047] **Figure 4.** An exemplary bispecific cloning strategy for integration within an expression enhancing locus. The three distinct antibody chains (AbC1, AbC2, and AbC3) of a bispecific antibody are first cloned into individual vectors. The AbC1 and AbC3 vectors each have RRS sites flanking the antibody expression cassette. Expression cassette for AbC2 is excised from the AbC2 plasmid and then subcloned into AbC1 expression plasmid, giving rise to a plasmid that contains, from 5' to 3', an RRS1 site, AbC1 expression cassette, AbC2 expression cassette, and an RRS3 site. This plasmid, together with the AbC3 plasmid and a recombinase, are introduced into host cells that harbor RRS1 and RRS2 in an expression enhancing locus. Bispecific antibody expression cell lines are isolated after recombinase-mediated cassette exchange

[0048] **Figure 5.** Expression of bispecific antibody from expression cassettes integrated at one genomic site (EESYR[®]). CHO cell lines RSX4189-1, RSX4187-1, RSX4191-1, RSX4188-1 were generated by recombinase-mediated cassette exchange at the EESYR[®] locus. The arrangements of expression cassettes for the three distinct antibody chains (AbC1, AbC2, and AbC3) of the bispecific Ab at the EESYR[®] locus are depicted on the left. The titers of each bispecific antibody in spent media of 4 day shaker flask cultures were determined by HPLC and are shown in the bar graph on the right.

DETAILED DESCRIPTION

Definitions

[0049] The term "antibody", as used herein, includes immunoglobulin molecules comprised of four polypeptide chains, two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds. Each heavy chain comprises a heavy chain variable region (abbreviated herein as HCVR or VH) and a heavy chain constant region. The heavy chain constant region comprises three domains, CH1, CH2 and CH3. Each light chain comprises a light chain variable region (abbreviated herein as LCVR or VL) and a light chain constant region. The light chain constant region comprises one domain, CL. The VH and VL regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each VH and VL is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4 (heavy chain CDRs may be abbreviated as HCDR1, HCDR2 and HCDR3; light chain CDRs may be abbreviated as LCDR1, LCDR2 and LCDR3).

[0050] The phrase "antigen-binding protein" includes a protein that has at least one CDR and is capable of selectively recognizing an antigen, i.e., is capable of binding an antigen with a K_D that is at least in the micromolar range. Therapeutic antigen-binding proteins (e.g., therapeutic antibodies) frequently require a K_D that is in the nanomolar or the picomolar range. Typically, an antigen-binding protein includes two or more CDRs, e.g., 2, 3, 4, 5, or 6 CDRs. Examples of antigen binding proteins include antibodies, antigen-binding fragments of antibodies such as polypeptides containing the variable regions of heavy chains and light

chains of an antibody (e.g., Fab fragment, F(ab')₂ fragment), and proteins containing the variable regions of heavy chains and light chains of an antibody and containing additional amino acids from the constant regions of heavy and/or light chains (such as one or more constant domains, i.e., one or more of CL, CH1, CH2, and CH3 domains).

[0051] The phrase "bispecific antigen-binding protein" includes antigen-binding proteins capable of selectively binding, or having different specificities to, two or more epitopes - either on two different molecules (e.g., antigens) or on the same molecule (e.g., on the same antigen). The antigen binding portion, or fragment antigen binding (Fab) portion of such protein renders specificity to a particular antigen, and is typically comprised of a heavy chain variable region and a light chain variable region of an immunoglobulin. In some circumstances, the heavy chain variable region and light chain variable region may not be a cognate pair, in other words, have a different binding specificities.

[0052] An example of a bispecific antigen-binding protein is a "bispecific antibody", which includes an antibody capable of selectively binding two or more epitopes. Bispecific antibodies generally comprise two different heavy chains, with each heavy chain specifically binding a different epitope - either on two different molecules (e.g., antigens) or on the same molecule (e.g., on the same antigen). If a bispecific antigen-binding protein is capable of selectively binding two different epitopes (a first epitope and a second epitope), the affinity of the variable region of the first heavy chain for the first epitope will generally be at least one to two or three or four orders of magnitude lower than the affinity of the variable region of the first heavy chain for the second epitope, and vice versa. Bispecific antigen-binding proteins such as bispecific antibodies can include the variable regions of heavy chains that recognize different epitopes of the same antigen. A typical bispecific antibody has two heavy chains each having three heavy chain CDRs, followed by (N-terminal to C-terminal) a CH1 domain, a hinge, a CH2 domain, and a CH3 domain, and an immunoglobulin light chain that either does not confer antigen-binding specificity but that can associate with each heavy chain, or that can associate with each heavy chain and that can bind one or more of the epitopes bound by the heavy chain antigen-binding regions, or that can associate with each heavy chain and enable binding of one or both of the heavy chains to one or both epitopes. In one

embodiment, an Fc domain includes at least CH2 and CH3. An Fc domain may include a hinge, a CH2 domain and CH3 domain.

[0053] One embodied bispecific format includes, a first heavy chain (HC), a second heavy chain which has a modified CH3 (HC*), and a common light chain (LC) (two copies of the same light chain). Another embodiment includes a first heavy chain (HC), a common LC and a HC-ScFv fusion polypeptide (wherein the second HC is fused to the N-terminus of the ScFv). Another embodiment includes a first HC, a cognate LC, an HC-ScFv fusion polypeptide (wherein the second HC is fused to the N-terminus of the ScFv). Another embodiment includes a first heavy chain (HC), a LC and an Fc domain. Another embodiment includes a first HC, an LC, an ScFv-Fc fusion polypeptide (wherein the Fc is fused to the C-terminus of the ScFv). Another embodiment includes a first HC, a common LC, and an Fc-ScFv fusion polypeptide (wherein the Fc is fused to the N-terminus of the ScFv). Another embodiment includes a first HC, a LC and an ScFv-HC (wherein the second HC is fused to the C-terminus of the ScFv).

[0054] In certain embodiments, one heavy chain (HC) may be native or "wild-type" sequence and the second heavy chain may be modified in the Fc domain. In other embodiments, one heavy chain (HC) may be native or "wild-type" sequence and the second heavy chain may be codon-modified.

[0055] The term "cell" includes any cell that is suitable for expressing a recombinant nucleic acid sequence, and has a locus that allows for stable integration and enhanced expression of an exogenous nucleic acid. Cells include mammalian cells, such as non-human animal cells, human cells, or cell fusions such as, for example, hybridomas or quadromas. In some embodiments, the cell is a human, monkey, ape, hamster, rat, or mouse cell. In some embodiments, the cell is a mammalian cell selected from the following cells: CHO (e.g., CHO K1, DXB-11 CHO, Veggie-CHO), COS (e.g., COS-7), retinal cell, Vero, CV1, kidney (e.g., HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK), HeLa, HepG2, WI38, MRC 5, Colo205, HB 8065, HL-60, (e.g., BHK21), Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L cell, C127 cell, SP2/0, NS-0, MMT 060562, Sertoli cell, BRL 3A cell, HT1080 cell, myeloma cell, tumor cell, and a cell line derived from an aforementioned cell. In some

embodiments, the cell comprises one or more viral genes, e.g. a retinal cell that expresses a viral gene (e.g., a PER.C6TM cell).

[0056] "Cell density" refers to the number of cells per volume of sample, for example as number of total (viable and dead) cells per mL. The number of cells may be counted manually or by automation, such as with a flow cytometer. Automated cell counters have been adapted to count the number of viable or dead or both viable/dead cells using for example a standard tryptan blue uptake technique. The phrase "viable cell density" or "viable cell concentration" refers to the number of viable cells per volume of sample (also referred to as "viable cell count"). Any number of well-known manual or automated techniques may be used to determine cell density. Online biomass measurements of the culture may be measured, where the capacitance or optical density is correlated to the number of cells per volume. Final cell density in a cell culture, such as in a production culture, varies depending on the starting cell line, for example in the range of about 1.0 to 10 x 10⁶ cells/mL. In some embodiments, final cell density reaches 1.0 to 10 x 10⁶ cells/mL prior to harvest of protein of interest from a production cell culture. In other embodiments, final cell density reaches at least 5.0 x 10⁶ cells/mL, at least 6 x 10⁶ cells/mL, at least 7 x 10⁶ cells/mL, at least 8 x 10⁶ cells/mL, at least 9 x 10⁶ cells/mL, or at least 10 x 10⁶ cells/mL.

[0057] The term "codon modified" means that a protein-coding nucleotide sequence has been modified in one or more nucleotides, i.e., one or more codons, without changing the amino acids encoded by the codons, resulting in a codon-modified version of the nucleotide sequence. Codon modification of a nucleotide sequence can provide a convenient basis to differentiate a nucleotide sequence from its codon-modified version in a nucleic acid-based assay (e.g., a hybridization based assays, PCR, among others). In some instances, codons of a nucleotide sequence are modified to provide improved or optimized expression of the encoded protein in a host cell by employing codon optimization techniques well known in the art (Gustafsson, C., et al., 2004, *Trends in Biotechnology*, 22:346–353; Chung, B.K.-S., et al., 2013, *Journal of Biotechnology*, 167:326–333; Gustafsson, C., et al., 2012, *Protein Expr Purif*, 83(1): 37–46). Sequence design software tools using such techniques are also well-known in the art, including but not limited to Codon optimizer (Fuglsang A. 2003, *Protein Expr Purif*, 31:247–249), Gene Designer (Villalobos A, et al., 2006, *BMC Bioinforma*, 7:285),

and OPTIMIZER (Puigbò P, et al. 2007, *Nucleic Acids Research*, 35:W126-W131), among others.

[0058] The phrase "complementarity determining region," or the term "CDR," includes an amino acid sequence encoded by a nucleic acid sequence of an organism's immunoglobulin genes that normally (i.e., in a wild-type animal) appears between two framework regions in a variable region of a light or a heavy chain of an immunoglobulin molecule (e.g., an antibody or a T cell receptor). A CDR can be encoded by, for example, a germline sequence or a rearranged or unrearranged sequence, and, for example, by a naive or a mature B cell or a T cell. In some circumstances (e.g., for a CDR3), CDRs can be encoded by two or more sequences (e.g., germline sequences) that are not contiguous (e.g., in an unrearranged nucleic acid sequence) but are contiguous in a B cell nucleic acid sequence, e.g., as the result of splicing or connecting the sequences (e.g., V-D-J recombination to form a heavy chain CDR3).

[0059] The term "expression enhancing locus" refers to a locus in the genome of a cell that contains a sequence or sequences and exhibits a higher level expression as compared to other regions or sequences in the genome when a suitable gene or construct is exogenously added (i.e., integrated) in or near the sequence or sequences, or "operably linked" to the sequence or sequences.

[0060] The term "enhanced" when used to describe enhanced expression includes an enhancement of at least about 1.5-fold to at least about 3-fold enhancement in expression over what is typically observed by random integration of an exogenous sequence into a genome or by integration at a different locus, for example, as compared to a pool of random integrants of a single copy of the same expression construct. Fold-expression enhancement observed employing the sequences of the invention is in comparison to an expression level of the same gene, measured under substantially the same conditions, in the absence of a sequence of the invention, for example in comparison to integration at another locus into the same species genome. Enhanced recombination efficiency includes an enhancement of the ability of a locus to recombine (for example, employing recombinase-recognition sites ("RRS")). Enhancement refers to an efficiency of recombination over random recombination for example, without employing recombinase-recognition sites or the like, which is typically

0.1%. A preferred enhanced recombination efficiency is about 10-fold over random, or about 1%. Unless specified, the claimed invention is not limited to a specific recombination efficiency. Enhanced expression loci typically support high productivity of the protein of interest by the host cell. Hence, enhanced expression includes high production of the protein of interest (elevated titer in grams of protein) per cell, rather than attaining high titers simply by high copy number of cells in culture. Specific productivity Q_p (pg/cell/day, i.e. pcd) is considered a measure of sustainable productivity. Recombinant host cells exhibiting Q_p greater than 5 pcd, or greater than 10 pcd, or greater than 15 pcd, or greater than 20 pcd, or greater than 25 pcd, or even greater than 30 pcd are desirable. Host cells with a gene of interest inserted into an expression-enhancing locus, or "hotspot", exhibit high specific productivity.

[0061] Where the phrase "exogenously added gene", "exogenously added nucleic acid", or simply "exogenous nucleic acid", is employed with reference to a locus of interest, the phrase refers to any DNA sequence or gene not present within the locus of interest as the locus is found in nature. For example, an "exogenous nucleic acid" within a CHO locus (e.g., a locus comprising a sequence of SEQ ID NO: 1 or SEQ ID NO: 2), can be a hamster gene not found within the particular CHO locus in nature (i.e., a hamster gene from another locus in the hamster genome), a gene from any other species (e.g., a human gene), a chimeric gene (e.g., human/mouse), or any other gene not found in nature to exist within the CHO locus of interest.

[0062] The phrase "heavy chain," or "immunoglobulin heavy chain" includes an immunoglobulin heavy chain constant region sequence from any organism, and unless otherwise specified includes a heavy chain variable domain. Heavy chain variable domains include three heavy chain CDRs and four FR regions, unless otherwise specified. A typical heavy chain has, following the variable domain (from N-terminal to C-terminal), a CH1 domain, a hinge, a CH2 domain, and a CH3 domain. The term "a fragment of a heavy chain" or "a heavy chain fragment" (also referred to herein as "HCF"), includes a peptide of at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more amino acids of a heavy chain, and may include one or more CDRs, one or more CDRs combined with one or more FRs, one or more of CH1, hinge, CH2, or CH3, the variable region, the constant region, fragments of the constant region

(e.g. CH1, CH2, CH3), or combinations thereof. Examples of an HCF include VHs, and full or parts of Fc regions. The phrase "a nucleotide sequence encoding an HCF" includes nucleotide sequences encoding a polypeptide consisting of an HCF and nucleotide sequences encoding a polypeptide containing an HCF, e.g., polypeptides that may contain additional amino acids in addition to a specified HCF. For example, a nucleotide sequence encoding an HCF includes nucleotide sequences encoding polypeptides consisting of a VH, consisting of a VH linked to a CH3, consisting of a full heavy chain, among others.

[0063] A "homologous sequence" in the context of nucleic acid sequences refers to a sequence that is substantially homologous to a reference nucleic acid sequence. In some embodiments, two sequences are considered to be substantially homologous if at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of their corresponding nucleotides are identical over a relevant stretch of residues. In some embodiments, the relevant stretch is a complete (i.e., full) sequence.

[0064] The phrase "light chain" includes an immunoglobulin light chain constant region sequence from any organism, and unless otherwise specified includes human kappa and lambda light chains. Light chain variable (VL) domains typically include three light chain CDRs and four framework (FR) regions, unless otherwise specified. Generally, a full-length light chain includes, from amino terminus to carboxyl terminus, a VL domain that includes FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4, and a light chain constant domain. Light chains that can be used with this invention include those, e.g., that do not selectively bind either the first or second epitope selectively bound by a bispecific antibody. Suitable light chains also include those that can bind or contribute to the binding of, one or both epitopes that are bound by the antigen-binding regions of an antibody. The term "a fragment of a light chain" or "a light chain fragment" (or "LCF") includes a peptide of at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more amino acids of a light chain, and may include one or more CDRs, one or more CDRs combined with one or more FRs, the variable region, the constant region, fragments of the constant region, or combinations thereof. Examples of an LCF include VLs and full or parts of light chain constant regions ("CLs"). The phrase "a nucleotide sequence encoding an LCF" includes nucleotide sequences encoding a polypeptide consisting of an LCF and nucleotide sequences encoding a polypeptide containing an LCF, e.g., polypeptides

that may contain additional amino acids in addition to a specified LCF. For example, a nucleotide sequence encoding an LCF includes nucleotide sequences encoding polypeptides consisting of a VL, or consisting of a full light chain, among others.

[0065] The phrase "operably linked" refers to linkage of nucleic acids or proteins in a manner that the linked molecules function as intended. DNA regions are operably linked when they are functionally related to each other. For example, a promoter is operably linked to a coding sequence if the promoter is capable of participating in the transcription of the sequence; a ribosome-binding site is operably linked to a coding sequence if it is positioned so as to permit translation. Generally, operably linked can include, but does not require, contiguity. In the case of sequences such as secretory leaders, contiguity and proper placement in a reading frame are typical features. An expression-enhancing sequence of the locus of interest is operably linked to a gene of interest (GOI) where it is functionally related to the GOI, for example, where its presence results in enhanced expression of the GOI.

[0066] "Percent identity", when describing a locus of interest, such as SEQ ID NO: 1 or SEQ ID NO: 2, or a fragment thereof, is meant to include homologous sequences that display the recited identity along regions of contiguous homology, but the presence of gaps, deletions, or insertions that have no homolog in the compared sequence are not taken into account in calculating percent identity.

[0067] As used herein, a "percent identity" determination between, *e.g.*, SEQ ID NO: 1, or fragment thereof, with a species homolog, would not include a comparison of sequences where the species homolog has no homologous sequence to compare in an alignment (*i.e.*, SEQ ID NO: 1 or the fragment thereof has an insertion at that point, or the species homolog has a gap or deletion, as the case may be). Thus, "percent identity" does not include penalties for gaps, deletions, and insertions.

[0068] "Recognition site" or "recognition sequence" is a specific DNA sequence recognized by a nuclease or other enzyme to bind and direct site-specific cleavage of the DNA backbone. Endonucleases cleave DNA within a DNA molecule. Recognition sites are also referred to in the art as recognition target sites.

[0069] "Recombinase recognition site" (or "RRS") is the specific DNA sequence recognized by a recombinase, such as Cre recombinase (Cre) or flippase (flp). Site-specific recombinases

can perform DNA rearrangements, including deletions, inversions and translocations when one or more of their target recognition sequences are placed strategically into the genome of an organism. In one example, Cre specifically mediates recombination events at its DNA target recognition site *loxP*, which is composed of two 13-bp inverted repeats separated by an 8-bp spacer. More than one recombinase recognition site may be employed, for example, to facilitate a recombination-mediated exchange of DNA. Variants or mutants of recombinase recognition sites, for example lox sites, may also be employed (Araki, N. et al, 2002, *Nucleic Acids Research*, 30:19, e103).

[0070] "Recombinase-mediated cassette exchange" or "RMCE" relates to a process for precisely replacing a genomic target cassette with a donor cassette. The molecular compositions typically provided in order to perform this process include 1) a genomic target cassette flanked both 5' and 3' by recognition target sites specific to a particular recombinase, 2) a donor cassette flanked by matching recognition target sites, and 3) the site-specific recombinase. Recombinase proteins are well known in the art (Turan, S. and Bode J., 2011, *FASEB J.*, 25, pp. 4088–4107) and enable precise cleavage of DNA within a specific recognition target site (sequence of DNA) without gain or loss of nucleotides. Common recombinase/site combinations include, but are not limited to, Cre/lox and Flp/ft. Commercially available kits also provide vectors containing the R4-*attP* site and a vector encoding the phiC31 integrase for RMCE. (See also, e.g. U.S. Published Application No. US20130004946.)

[0071] "Site-specific integration" or "targeted insertion" refers to gene targeting methods employed to direct insertion or integration of a gene or nucleic acid sequence to a specific location in the genome, *i.e.*, to direct the DNA to a specific site between two nucleotides in a contiguous polynucleotide chain. Site-specific integration or targeted insertion may also be done for a particular nucleic acid that includes multiple expression units or cassettes, such as multiple genes, each having their own regulatory elements (such as promoters, enhancers, and/or transcriptional termination sequences). "Insertion" and "integration" are used interchangeably. It is understood that insertion of a gene or nucleic acid sequence (for example a nucleic acid sequence comprising an expression cassette) may result in (or may be

engineered for) the replacement or deletion of one or more nucleic acids depending on the gene editing technique being utilized.

[0072] "Stable integration" means that an exogenous nucleic acid integrated in the genome of a host cell remains integrated for an extended period of time in cell culture, for example, at least 7 days, at least 10 days, at least 15 days, at least 20 days, at least 25 days, at least 30 days, at least 35 days, at least 40 days, at least 45 days, at least 50 days, at least 55 days, at least 60 days, or longer. It is understood that making bispecific antigen-binding proteins for manufacturing and purification at large-scale is a challenging task. Stability and clonality are essential to the reproducibility of any biomolecule, especially one to be used therapeutically. The stable clones expressing bispecific antibodies made by the methods of this disclosure provide a consistent and reproducible way to generate therapeutic biomolecules.

General Description

[0073] This disclosure provides for compositions and methods for improved expression of multiple polypeptides in a host cell particularly Chinese hamster (*Cricetulus griseus*) cell lines, by employing an expression-enhancing locus in the host cell. More specifically, the disclosure provides compositions and methods designed to integrate multiple exogenous nucleic acids that together encode a bispecific antigen-binding protein into a specific site within an expression-enhancing locus in a host cell such as a CHO cell. In particular, this disclosure provides cells containing multiple exogenous nucleic acids integrated at a specific site within an expression-enhancing locus wherein the multiple exogenous nucleic acids together encode a bispecific antigen-binding protein. This disclosure further provides nucleic acid vectors designed for site-specific integration of multiple exogenous nucleic acids into an expression-enhancing locus. This disclosure additionally provides systems that include a host cell containing two or more recombinase recognition sites (RRSs), and a set of vectors containing matching RRSs and multiple exogenous nucleic acids, for site-specific integration of the multiple exogenous nucleic acids from the vectors into an expression-enhancing locus. Further, this disclosure provides methods for making a bispecific antigen-binding protein using the cells, vectors and systems disclosed herein.

Cells Having Multiple Exogenous Nucleic Acids Integrated at a Specific Site within an Expression Enhancing Locus

[0074] In one aspect, this disclosure provides a cell containing an exogenous nucleic acid sequence integrated at a specific site within an enhanced expression locus, wherein the exogenous nucleic acid sequence encodes a bispecific antigen-binding protein.

[0075] The cells provided herein are capable of producing a bispecific antigen-binding protein (e.g., a bispecific antibody) with high titers and/or high specific productivity (pg/cell/day). In some embodiments, a cell produces a bispecific antigen-binding protein at a titer of at least 5 mg/L, 10 mg/L, 15 mg/L, 20 mg/L, 25 mg/L, 30 mg/L, 35 mg/L, 40 mg/L, 45 mg/L, 50 mg/L, or greater. In some embodiments, a cell produces a bispecific antigen-binding protein at a ratio of the bispecific antigen-binding protein titer versus the total antigen-binding protein titer of at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, or higher. In some embodiments, a cell that produces a bispecific antigen-binding protein has a specific productivity of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 picogram/cell/day (pcd), or higher, determined based on total antigen-binding proteins (in pg) produced per cell per day.

[0076] The host cells comprising an exogenous nucleic acid sequence encoding a bispecific antigen-binding protein integrated at a specific site within an enhanced expression locus exhibit high cell density in production culture, e.g. 1 to 10×10^6 cells/mL. In other embodiments, the bispecific antigen-binding protein-encoding host cell has a final cell density of at least 5×10^6 cells/mL, 6×10^6 cells/mL, 7×10^6 cells/mL, 8×10^6 cells/mL, 9×10^6 cells/mL, or 10×10^6 cells/mL.

[0077] In some embodiments, the bispecific antigen-binding protein contains two HC fragments ("HCFs") having different antigen specificities and two LCFs. In instances where two VL regions are used, they can be the same or different. In specific embodiments, the two VL regions are the same, such as a common light chain.

[0078] In some embodiments, each of the two HCFs includes amino acids from a heavy chain constant region, such as CH1, CH2, or CH3. In specific embodiments, each of the two HCFs includes a CH3 domain. In particular embodiments, each of two HCFs include a constant region, i.e., a full constant region.

[0079] In some embodiments, each of the two HCFs include a VH, and the two VHs can be the same or different.

[0080] In some embodiments, the bispecific antigen-binding protein includes two heavy chains (i.e., two full heavy chains).

[0081] In some embodiments, each of the two LCFs includes a VL. In specific embodiments, each LCF consists of a VL region that is operably linked to an amino acid sequence that includes amino acids from a light chain constant region. In specific embodiments, each VL region is operably linked to a CL region, i.e., the bispecific antigen-binding protein includes a light chain (i.e., a full light chain).

[0082] In some embodiments, the exogenous nucleic acid sequence integrated within an enhanced expression locus includes a first exogenous nucleic acid containing a nucleotide sequence encoding a first LCF, a second exogenous nucleic acid containing a nucleotide sequence encoding a first HCF, and a third exogenous nucleic acid containing a nucleotide sequence encoding a second HCF.

[0083] In some embodiments, the nucleotide sequence encoding the first LCF can encode a light chain variable (VL) region sequence. In specific embodiments, the nucleotide sequence encoding the first VL region encode a first light chain.

[0084] In some embodiments, the nucleotide sequence encoding the first HCF encodes amino acids from a first heavy chain constant region, e.g., one or more of CH1, hinge, CH2, or CH3), and the nucleotide sequence encoding the second HCF encodes amino acids from a second heavy chain constant region. The amino acids from a first heavy chain constant region can be the same or different from the amino acids from a second heavy chain constant region. For example, the nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain, wherein the first and second CH3 domains can be the same, or different in one or more amino acid positions, as described herein below for bispecific antigen-binding proteins.

[0085] In some embodiments, the nucleotide sequence encoding the first HCF encodes a first VH, and the nucleotide sequence encoding the second HCF encodes a second VH.

[0086] In some embodiments, the nucleotide sequence encoding the first HCF encodes a first heavy chain, and the nucleotide sequence encoding the second HCF encodes a second heavy

chain. The first and second heavy chains can have the same constant regions, or differ in one or more amino acids. Various examples of bispecific antigen-binding proteins that having different heavy chain constant domains (such as different CH3 domains) are further described herein below. Independent of the encoded amino acid sequences, the nucleotide sequences encoding amino acids from two heavy chain constant regions can differ in that one of the two coding nucleotide sequences can be codon modified, which provides a convenient basis to differentiate the two nucleotide sequences in a nucleic acid-based detection assay.

[0087] In some embodiments, each HCF- or LCF-coding nucleotide sequence is independently and operably linked to a transcriptional regulatory sequence that contains a promoter. By "independently", it means that each coding sequence is operably linked to a separate transcriptional regulatory sequence such as a promoter, so that the transcription of the coding sequences is under separate regulation and control. In some embodiments, the promoters directing transcription of the two HCF-containing polypeptides are the same. In some embodiments, the promoters directing transcription of the two HCF-containing polypeptides, as well as the promoter directing transcription of the LCF-containing polypeptide, are all the same, e.g., a CMV promoter. In some embodiments, each HCF- or LCF-coding nucleotide sequence is independently and operably linked to an inducible or repressible promoter. Inducible or repressible promoters allow production to occur, for example, only in production phase (fed-batch culture) and not during growth phase (seed train culture). Fine control of production (expression) of each gene product may be achieved by way of different promoters.

[0088] In one such example, cells are first engineered to express the tetracycline repressor protein (TetR) and each HCF- and LCF-coding nucleotide sequence is placed under transcriptional control of a promoter whose activity is regulated by TetR. Two tandem TetR operators (TetO) are placed immediately downstream of the CMV promoter. In some embodiments, each HCF- and/or LCF-coding nucleotide sequence is independently and operably linked to a promoter upstream of at least one TetR operator (TetO) or Arc operator (ArcO). In other embodiments, each HCF- and/or LCF-coding nucleotide sequence is independently and operably linked to a CMV/TetO or CMV/ArcO hybrid promoter. Additional suitable promoters are described herein below.

[0089] The relative positions of the multiple exogenous nucleic acids within the locus can vary. Without intending to be bound by any theory, it is believed that it is important to achieve balanced (i.e., comparable) expression levels of the two HCF-containing polypeptides. In some embodiments, the LCF encoding nucleic acid is located upstream relative to both HCF-encoding nucleic acids. In instances where the three promoters for directing the expression of the LCF-containing polypeptide and the two HCF-containing polypeptides are the same, a suitable arrangement can include, from 5' to 3', the nucleotide sequence encoding the LCF, the nucleotide sequence encoding the first HCF, an additional different promoter operably linked to a nucleotide sequence (such as a selectable marker gene), and the nucleotide sequence encoding the second HCF. Other suitable arrangements include, from 5' to 3', the nucleotide sequence encoding the LCF, an additional different promoter operably linked to a nucleotide sequence (such as a selectable marker gene), the nucleotide sequence encoding the first HCF, and the nucleotide sequence encoding the second HCF. Where the nucleotide sequences encoding a HCF encodes a constant region sequence, either the nucleotide sequence located upstream can encode a modified version of a constant region sequence (e.g., a modified CH3), or the nucleotide sequence located upstream can encode a modified version of a constant region sequence, with the other one encoding the unmodified version of the constant region sequence.

[0090] In some embodiments, the cell further contains one or more RRS, also integrated within the locus. In some embodiments, the cell includes a first and a second RRS different from each other and flanking an exogenous nucleic acid sequence, wherein the exogenous nucleic acid sequence in turn contains a first LCF-encoding nucleic acid, a first HCF-encoding nucleic acid, and a second HCF-encoding nucleic acid. In specific embodiments, the LCF encoding nucleic acid is located upstream relative to both HCF-encoding nucleic acids, and the cell includes a third RRS, located 3' relative to the first LCF-encoding nucleic acid, and 5' relative to one or both of the HCF-encoding exogenous nucleic acids, wherein the third RRS is different from the first and second RRSs. The third RRS can be engineered to be included in an intron of a gene which can be placed between any two of the HCF- or LCF-encoding sequences.

Bispecific Antigen-Binding Proteins

[0091] Bispecific antigen-binding proteins, such as bispecific antibodies, suitable for cloning and production in the cells, vectors, and systems described in this disclosure are not limited to any particular format of bispecific antigen-binding proteins.

[0092] In various embodiments, the bispecific antigen-binding protein includes two polypeptides, each containing an antigen-binding moiety (e.g., a VH region) and a CH3 domain, wherein the antigen-binding moiety of the two polypeptides have different antigen specificities, and wherein the two CH3 domains are heterodimeric in respect to each other in that one of the CH3 domains has been modified in at least one amino acid position to give rise to differential Protein A binding characteristics between the two polypeptides. See, e.g., the bispecific antibodies described in U.S. Patent 8,586,713. In this way, a differential protein A isolation scheme can be employed to readily isolate the heterodimeric bispecific antigen-binding proteins from homodimers.

[0093] In some embodiments, the bispecific antigen-binding protein includes two heavy chains having different antigen specificities and differing in at least one amino acid position in the CH3 domain to give rise to differential Protein A binding characteristics between the two heavy chains.

[0094] In some embodiments, the two polypeptides contain CH3 domains of human IgG, wherein one of the two polypeptides contains the CH3 domain of a human IgG selected from IgG1, IgG2 and IgG4, and the other one of the two polypeptides contains a modified CH3 domain of a human IgG selected from IgG1, IgG2 and IgG4 wherein the modification reduces or eliminates the binding of the modified CH3 region to Protein A. In specific embodiments, one of the two polypeptides contains the CH3 domain of human IgG1, and the other one of the two polypeptides contains a modified CH3 domain of human IgG1 wherein the modification is selected from the group consisting of (i) 95R and (ii) 95R and 96F in the IMGT exon numbering system. In other specific embodiments, the modified CH3 domain comprises one to five additional modifications selected from the group consisting of 16E, 18M, 44S, 52N, 57M, and 82I in the IMGT exon numbering system.

[0095] In other various embodiments, the two polypeptides contain CH3 domains of mouse IgG, wherein one of the two polypeptides contains the CH3 domain of an unmodified mouse

IgG, and the other one of the two polypeptides contains a modified CH3 domain of the mouse IgG wherein the modification reduces or eliminates the binding of the modified CH3 region to Protein A. In various embodiments, a mouse IgG CH3 region is modified to comprise particular amino acids at particular positions (EU numbering), selected from the group consisting of: 252T, 254T, and 256T; 252T, 254T, 256T, and 258K; 247P, 252T, 254T, 256T, and 258K; 435R and 436F; 252T, 254T, 256T, 435R, and 436F; 252T, 254T, 256T, 258K, 435R, and 436F; 247P, 252T, 254T, 256T, 258K, 435R, and 436F; and, 435R. In a specific embodiment, a particular group of modifications is made, selected from the groups consisting of: M252T, S254T, S256T; M252T, S254T, S256T, I258K; I247P, M252T, S254T, S256T, I258K; H435R, H436F; M252T, S254T, S256T, H435R, H436F; M252T, S254T, S256T, I258K, H435R, H436F; I247P, M252T, S254T, S256T, I258K, H435R, H436F; and, H435R.

[0096] In various embodiments, a bispecific antigen-binding protein is a hybrid of a mouse and a rat monoclonal antibody or antigen-binding protein, e.g., a hybrid of mouse IgG2a and rat IgG2b. According to these embodiments, a bispecific antibody is composed of a heterodimer of the two antibodies comprising one heavy/light chain pair of each, associating via their Fc portions. The desired heterodimer can be easily purified from a mixture of two parental antibody homodimers and the bispecific heterodimer, because the binding properties of the bispecific antibody to Protein A are different from those of the parental antibodies: rat IgG2b does not bind to protein A, whereas the mouse IgG2a does. Consequently, the mouse-rat heterodimer binds to Protein A but elutes at a higher pH than the mouse IgG2a homodimer, and this makes selective purification of the bispecific heterodimer possible.

[0097] In other various embodiments, a bispecific antigen-binding protein is of a format that is referred to as "knobs-into-holes" in the art (see, e.g., U.S. Patent No. 7,183,076). In these embodiments, the Fc portions of two antibodies are engineered to give one a protruding "knob", and the other a complementary "hole." When produced in the same cell, the heavy chains are said to preferentially form heterodimers rather than homodimers, by association of the engineered "knobs" with the engineered "holes."

[0098] In another embodiment, the first heavy chain and the second heavy chain comprises one or more amino acid modifications in the CH3 domain to enable interaction between two heavy chains. CH3-CH3 interface amino acid residues can be replaced with charged amino

acid to provide electrostatically unfavorable homodimer formation. (See, e.g. PCT Publication No. WO2009089004; and European Publication No. EP1870459.)

[0099] In other embodiments, the first heavy chain comprises a CH3 domain of the isotype IgA and the second heavy chain comprises a CH3 domain of IgG (or vice versa) to promote preferential formation of heterodimers. (See e.g. PCT Publication No. WO2007110205.)

[00100] In other embodiments, various formats can be incorporated with immunoglobulin chains by engineering methods to foster formation of heterodimers, such as Fab-arm exchange (PCT Publication No. PCT Publication No. WO2008119353; PCT Publication No.

WO2011131746), coiled-coil domain interaction (PCT Publication No. WO2011034605) or leucine zipper peptides (Kostelnik, et al. *J. Immunol.* 1992, 148(5):1547-1553).

[00101] Immunoglobulin heavy chain variable regions that can be used to generate bispecific antigen-binding proteins can be generated using any method known in the art. For example, a first heavy chain comprises a variable region that is encoded by a nucleic acid that is derived from the genome of a mature B cell of a first animal that has been immunized with a first antigen, and the first heavy chain specifically recognizes the first antigen; and a second heavy chain comprises a variable region that is encoded by a nucleic acid that is derived from the genome of a mature B cell of a second animal that has been immunized with a second antigen, and the second heavy chain specifically recognizes the second antigen. Immunoglobulin heavy chain variable region sequences can also be obtained by any other method known in the art, e.g., by phage display. In other examples, nucleic acids encoding the heavy chain variable regions include those of antibodies that have been described or otherwise available in the art. In some embodiments, one of the two heavy chain coding sequences have been codon modified in order to provide a convenient basis to differentiate the two coding sequences in nucleic acid based assays.

[00102] Bispecific antibodies comprising two heavy chains that recognize two different epitopes (or two different antigens) are more easily isolated where they can pair with the same light chain (i.e., light chains having identical variable and constant domains). A variety of methods are known in the art for generating light chains that can pair with two heavy chains of differing specificity, while not interfering or not substantially interfering with the

selectivity and/or affinity of the heavy chain variable domain with its target antigen, as described in e.g., U.S. Patent 8,586,713 and the art disclosed therein.

[00103] The bispecific antigen-binding proteins can have a variety of dual antigen specificities and associated useful applications.

[00104] In some examples, bispecific antigen-binding proteins that comprise binding specificity toward a tumor antigen and a T-cell antigen can be made that target an antigen on a cell, e.g., CD20, and also target an antigen on a T-cell, e.g., a T cell receptor such as CD3. In this way, the bispecific antigen-binding protein targets both a cell of interest in a patient (e.g., B cell in a lymphoma patient, via CD20 binding) as well as a T-cell of the patient. The bispecific antigen-binding protein, in various embodiments, is designed so as to activate the T-cell upon binding a T cell receptor such as binding to CD3, thus coupling T-cell activation to a specific, selected tumor cell.

[00105] In the context of bispecific antigen-binding proteins wherein one moiety binds to CD3 and the other moiety binds to a target antigen, the target antigen can be a tumor-associated antigen. Non-limiting examples of specific tumor-associated antigens include, e.g., AFP, ALK, BAGE proteins, BIRC5 (survivin), BIRC7, β -catenin, brc-abl, BRCA1, BCMA, BORIS, CA9, carbonic anhydrase IX, caspase-8, CALR, CCR5, CD19, CD20(MS4A1), CD22, CD30, CD40, CDK4, CEA, CLEC-12, CTLA4, cyclin-B1, CYP1B1, EGFR, EGFRvIII, ErbB2/Her2, ErbB3, ErbB4, ETV6-AML, EpCAM, EphA2, Fra-1, FOLR1, GAGE proteins (e.g., GAGE-1, -2), GD2, GD3, GloboH, glypican-3, GM3, gp100, Her2, HLA/B-raf, HLA/k-ras, HLA/MAGE-A3, hTERT, LMP2, MAGE proteins (e.g., MAGE-1, -2, -3, -4, -6, and -12), MART-1, mesothelin, ML-IAP, Muc1, Muc2, Muc3, Muc4, Muc5, Muc16 (CA-125), MUM1, NA17, NY-BR1, NY-BR62, NY-BR85, NY-ESO1, OX40, p15, p53, PAP, PAX3, PAX5, PCTA-1, PLAC1, PRLR, PRAME, PSMA (FOLH1), RAGE proteins, Ras, RGS5, Rho, SART-1, SART-3, Steap-1, Steap-2, TAG-72, TGF- β , TMPRSS2, Thompson-nouvelle antigen (Tn), TRP-1, TRP-2, tyrosinase, and uroplakin-3.

[00106] In some embodiments, the bispecific antigen-binding protein is selected from the group consisting of an anti-CD3 x anti-CD20 bispecific antibody (as described in U.S. Pat. Appln. Pub. Nos. US2014/0088295A1 and US20150266966A1, herein incorporated by reference), an anti-CD3 x anti-Mucin 16 bispecific antibody (e.g., an anti-CD3 x anti-Muc16

bispecific antibody), and an anti-CD3 x anti- Prostate-specific membrane antigen bispecific antibody (e.g., an anti-CD3 x anti-PSMA bispecific antibody). In other embodiments, the bispecific antigen-binding protein comprises one moiety that binds CD3. Exemplified anti-CD3 antibody moieties are described in U.S. Pat. Appln. Pub. Nos. US2014/0088295A1 and US2015/0266966A1, and in International Publication No. WO 2017/053856 published on March 30, 2017, all of which are incorporated herein by reference). In still other embodiments, the bispecific antigen-binding protein comprises one moiety that binds to CD3 and one moiety that binds to BCMA, CD19, CD20, CD28, CLEC-12, Her2, HLA protein, MAGE protein, Muc16, PSMA, or Steap-2.

[00107] In the context of bispecific antigen-binding proteins wherein one moiety binds to a T cell receptor such as binds to CD3 and the other moiety binds a target antigen, the target antigen can be an infectious disease-associated antigen. Non-limiting examples of infectious disease-associated antigens include, *e.g.*, an antigen that is expressed on the surface of a virus particle, or preferentially expressed on a cell that is infected with a virus, wherein the virus is selected from the group consisting of HIV, hepatitis (A, B or C), herpes virus (*e.g.*, HSV-1, HSV-2, CMV, HAV-6, VZV, Epstein Barr virus), adenovirus, influenza virus, flavivirus, echovirus, rhinovirus, coxsackie virus, coronavirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV, dengue virus, papillomavirus, molluscum virus, poliovirus, rabies virus, JC virus, and arboviral encephalitis virus. Alternatively, the target antigen can be an antigen that is expressed on the surface of a bacterium, or preferentially expressed on a cell that is infected with a bacterium, wherein the bacterium is selected from the group consisting of chlamydia, rickettsia, mycobacteria, staphylococci, streptococci, pneumonococci, meningococci, gonococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospira, and Lyme disease bacteria. In certain embodiments, the target antigen is an antigen that is expressed on the surface of a fungus, or preferentially expressed on a cell that is infected with a fungus, wherein the fungus is selected from the group consisting of *Candida* (*albicans*, *krusei*, *glabrata*, *tropicalis*, etc.), *Cryptococcus neoformans*, *Aspergillus* (*fumigatus*, *niger*, etc.), *Mucorales* (*mucor*, *absidia*, *rhizopus*, etc.), *Sporothrix schenckii*, *Blastomyces dermatitidis*, *Paracoccidioides brasiliensis*, *Coccidioides immitis*, and

Histoplasma capsulatum. In certain embodiments, the target antigen is an antigen that is expressed on the surface of a parasite, or preferentially expressed on a cell that is infected with a parasite, wherein the parasite is selected from the group consisting of *Entamoeba histolytica*, *Balantidium coli*, *Naegleria fowleri*, *Acanthamoeba* sp., *Giardia lamblia*, *Cryptosporidium* sp., *Pneumocystis carinii*, *Plasmodium vivax*, *Babesia microti*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, *Nippostrongylus brasiliensis*, *Taenia crassiceps*, and *Brugia malayi*. Non-limiting examples of specific pathogen-associated antigens include, e.g., HIV gp120, HIV CD4, hepatitis B glucoprotein L, hepatitis B glucoprotein M, hepatitis B glucoprotein S, hepatitis C E1, hepatitis C E2, hepatocyte-specific protein, herpes simplex virus gB, cytomegalovirus gB, and HTLV envelope protein.

[00108] Bispecific binding proteins that comprise two binding moieties that are each directed to a binding partner (i.e., each directed to a different target) on the surface of the same cell can also be made. This design is particularly suited to targeting specific cells or cell types that express both targets on the surface of the same cell. Although targets might appear individually on other cells, the binding moieties of these binding proteins are selected such that each binding moiety binds its target with a relatively low affinity (e.g., low micromolar, or high nanomolar - e.g., over a hundred nanomolar KD, e.g., 500, 600, 700, 800 nanomolar). In this way, prolonged target binding is favored only in situations where the two targets are in proximity on the same cell.

[00109] Bispecific binding proteins that comprise two binding moieties that bind the same target, each at a different epitope of the same target, can be made. This design is particularly suited for maximizing the probability of successfully blocking a target with binding protein. Multiple extracellular loops, e.g., of a transmembrane channel or a cell surface receptor, can be targeted by the same bispecific binding molecule.

[00110] Bispecific binding proteins that comprise two binding moieties that cluster and activate negative regulators of immune signaling to result in immune suppression can be made. Repression in *cis* can be achieved where the targets are on the same cell; repression in *trans* can be achieved where the targets are on different cells. Repression in *cis*, e.g., can be achieved with a bispecific binding protein having an anti-IgGRIIb binding moiety and an anti-

Fc ϵ RI binding moiety, such that the IgGRIIb is clustered only in the presence of Fc ϵ RI, in order to down-regulate an immune response to Fc ϵ RI. Repression in trans, e.g., can be achieved with a bispecific binding protein having an anti-BTLA binding moiety and a binding moiety that specifically binds a tissue-specific antigen of interest, such that clustering of the inhibitory BTLA molecule occurs only in the selected target tissue, which potentially addresses auto-immune diseases.

[00111] Bispecific binding proteins that activate multi-component receptors can be made. In this design, two binding moieties directed to two components of a receptor bind, cross-link the receptor, and activate signaling from the receptor. This can be done, e.g., using a bispecific binding protein with a binding moiety that binds IFNAR1 and a binding moiety that binds IFNAR2, where binding cross-links the receptor. Such a bispecific binding protein can provide an alternative to interferon treatment.

[00112] Bispecific binding proteins that transport binding moieties across a semi-permeable barrier, e.g., the blood-brain barrier, can be made. In this design, one binding moiety binds a target that can transit a particular selective barrier; the other binding moiety targets a molecule with a therapeutic activity, wherein the target molecule with therapeutic activity cannot normally traverse the barrier. This kind of bispecific binding protein is useful for bringing therapeutics to tissues that the therapeutic would not otherwise reach. Some examples include targeting the pIGR receptor to transport a therapeutic into the gut or lung, or targeting the transferrin receptor to transport a therapeutic across the blood-brain barrier.

[00113] Bispecific binding proteins that transport binding moieties into specific cells or cell types can be made. In this design, one binding moiety targets a cell surface protein (e.g., a receptor) that is readily internalized into the cell. The other binding moiety targets an intracellular protein, where binding of the intracellular protein results in a therapeutic effect.

[00114] Bispecific binding proteins that bind a surface receptor of a phagocytic immune cell and a surface molecule of an infectious pathogen (e.g., a yeast or bacterium), to bring the infectious pathogen in the vicinity of a phagocytic immune cell to facilitate phagocytosis of the pathogen. An example of such a design would be a bispecific antibody that targets a CD64 or CD89 molecule and also a pathogen.

[00115] Bispecific binding proteins that have an antibody variable region as one binding moiety and a non-Ig moiety as a second binding moiety. The antibody variable region achieves targeting, whereas the non-Ig moiety is an effector or a toxin linked to an Fc. In this way, the ligand (e.g., an effector or toxin) is delivered to the target bound by the antibody variable region.

[00116] Bispecific binding proteins that have two moieties each bound to an Ig region (e.g., an Ig sequence containing a CH2 and CH3 region) such that any two protein moieties can be brought in each other's vicinity in the context of the Fc. Examples of this design include traps, e.g., homo- or heterodimeric trap molecules.

Expression-Enhancing Loci

[00117] Expression-enhancing loci suitable for use in this invention include for example, a locus that comprises a nucleotide sequence having substantial homology to SEQ ID NO: 1 as described in U.S. Patent 8,389,239 (also referred to herein as the "EESYR® locus"), a locus that comprises a nucleotide sequence having substantial homology to SEQ ID NO: 2 or SEQ ID NO: 3 as described in U.S. Application Serial No. 14/919,300 (also referred to herein as "the YARS locus"), and other expression-enhancing loci and sequences documented in the art (e.g., US 20150167020A1, and U.S. Patent 6,800,457).

[00118] In some embodiments, the expression-enhancing locus that is used in this invention is selected from a locus that comprises a nucleotide sequence having substantial homology to SEQ ID NO: 1, or a locus that comprises a nucleotide sequence having substantial homology to SEQ ID NO: 2 or SEQ ID NO: 3. These loci contain sequences that not only provide for enhanced expression of genes integrated in operable linkage to the sequences (i.e., within the sequences or within close proximity to the sequences), but also exhibit greater recombination efficiency and improved integration stability, as compared to other sequences in the genome.

[00119] SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3 have been identified from CHO cells. Other mammalian species (such as, for example, humans or mice), were found to have limited homology to the identified expression-enhancing region; however, homologous sequences may be found in cell lines derived from other tissue types of *Cricetulus griseus*, or other homologous species, and can be isolated by techniques that are well-known in the art. For example, one may identify other homologous sequences by cross-species hybridization or

PCR-based techniques. In addition, changes can be made in the nucleotide sequence set forth in SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3 by site-directed or random mutagenesis techniques that are well known in the art. The resulting sequence variants can then be tested for expression-enhancing activity. DNAs that are at least about 90% identical in nucleic acid identity to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3 having expression-enhancing activity are isolatable by routine experimentation, and are expected to exhibit expression-enhancing activity.

[00120] The integration site, the site or nucleotide position of insertion of one or more exogenous nucleic acids, can be at any position that is within or adjacent to any of the expression enhancing sequences (such as SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3). Whether a specific chromosomal location within or adjacent to the locus of interest supports stable integration and efficient transcription of an integrated exogenous gene can be determined in accordance with standard procedures well known in the art, e.g., as described U.S. Patent 8,389,239 and U.S. Application Serial No. 14,919,300.

[00121] The integration sites considered herein are located within the expression enhancing sequences, or within close proximity to the sequences, e.g., less than about 1 kb, 500 base pairs (bp), 250 bp, 100 bp, 50 bp, 25 bp, 10 bp, or less than about 5 bp upstream (5') or downstream (3') with respect to the location of an expression enhancing sequence on the chromosomal DNA. In still some other embodiments, the employed integration site is located at about 1000, 2500, 5000 or more base pairs upstream (5') or downstream (3') with respect to the location of an expression enhancing sequence on the chromosomal DNA.

[00122] It is understood in the art that large genomic regions, such as scaffold/matrix attachment regions, are employed for efficient replication and transcription of chromosomal DNA. A scaffold/matrix attachment region (S/MAR), also known as called scaffold-attachment region (SAR), or matrix-associated or matrix attachment region (MAR), is a eukaryotic genomic DNA region where the nuclear matrix attaches. Without being bound by any one theory, S/MARs typically map to non-coding regions, separate a given transcriptional region (e.g. chromatin domain) from its neighbors, and also provide platforms for the machinery and/or binding of factors that enable transcription, such as recognition sites for DNases or polymerases. Some S/MARs have been characterized at about 14-20 kb in length

(Klar, et al. 2005, *Gene* 364:79-89). As such, integration of genes at an expression enhancing locus (e.g., within or near SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO:3) is expected to confer enhanced expression. In some embodiments, the host cells comprising an exogenous nucleic acid sequence encoding a bispecific antigen-binding protein integrated at a specific site within an enhanced expression locus exhibits high specific productivity. In other embodiments, the bispecific antigen-binding protein-encoding host cell has a specific productivity of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, or 30 picogram/cell/day (pcd).

[00123] In some embodiments, the integration site is within a locus that comprises the nucleotide sequence of SEQ ID NO: 1. In specific embodiments, the integration site is within, or within close proximity to, the nucleotide sequence of SEQ ID NO: 1. In particular embodiments, the integration site is at a position within SEQ ID NO: 1 selected from nucleotides spanning positions numbered 10-13,515; 20-12,020; 1,020-11,020; 2,020-10,020; 3,020-9,020; 4,020-8,020; 5,020-7,020; 6,020-6,920; 6,120-6,820; 6,220-6,720; 6,320-6,620; 6,420-6,520; 6,460-6,500; 6,470-6,490; and 6,475-6,485. In other embodiments, the integration site is in a sequence that is selected from the group consisting of nucleotides 5,000-7,400, 5,000-6,500, 6,400-7,400 of SEQ ID NO: 1; and nucleotides 6,400-6,500 of SEQ ID NO: 1. In a specific embodiment, the integration site before, after, or within the "act" triplet of nucleotides 6471 to 6473 of SEQ ID NO: 1.

[00124] In some embodiments, the integration site is within a locus that comprises the nucleotide sequence of SEQ ID NO: 2 or SEQ ID NO:3. In specific embodiments, the integration site is within, or within close proximity to, the nucleotide sequence of SEQ ID NO: 2. In particular embodiments, the integration site is within, or within close proximity to, the nucleotide sequence of SEQ ID NO: 3. In some embodiments, the integration site is within nucleotides 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1995-1996, 1996-1997, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021 or 2021-2022 of SEQ ID NO: 3. In specific embodiments, the integration is at or within nucleotides 2001-2022 of SEQ ID NO: 3. In some embodiments, the exogenous nucleic acid is inserted at

or within nucleotides 2001-2002 or nucleotides 2021-2022 of SEQ ID NO: 3 and nucleotides 2002-2021 of SEQ ID NO: 3 are deleted, as a result of the insertion.

Site-Specific Integration Into An Expression-Enhancing Locus

[00125] Integration of multiple exogenous nucleic acids into an expression-enhancing locus in a site-specific manner, i.e., into one specific site within an expression-enhancing locus as disclosed herein, can be achieved in several ways including, e.g., by homologous recombination, and recombinase mediated cassette exchange, as described in the art (see e.g., U.S. Patent 8,389,239 and the art disclosed therein).

[00126] In some embodiments, cells are provided that contain at least two, i.e., two or more, different recombinase recognition sequences (RRS) within an expression-enhancing locus convenient for integrating an nucleic acid sequence containing multiple exogenous nucleic acids or genes of interest. Such cells can be obtained by introducing an exogenous nucleic acid sequence containing two or more RRS into a desirable locus by various means including homologous recombination, as described hereinbelow and in the art, e.g., U.S. Patent 8,389,239 and the art disclosed therein.

[00127] In specific embodiments, cells are provided that contain more than two different recombinase recognition sequences (RRS) within an expression-enhancing locus convenient for integrating multiple exogenous nucleic acids. In particular embodiments, cells are provided that contain three different recombinase recognition sequences (RRS) within an expression-enhancing locus which can mediate integration of two separate exogenous nucleic acids, for example, wherein the 5' RRS and the middle RRS in the genome match the 5' RRS and the 3' RRS flanking the first exogenous nucleic acid to be integrated, and the middle RRS and 3' RRS in the genome match the 5' RRS and the 3' RRS flanking the second exogenous nucleic acid to be integrated.

[00128] Suitable RRSs can be selected from the group comprising *LoxP*, *Lox511*, *Lox5171*, *Lox2272*, *Lox2372*, *Loxm2*, *Lox-FAS*, *Lox71*, *Lox66* and the mutants thereof, where the site specific recombinase is Cre recombinase or its derivative is used to achieve recombinase-mediated cassette exchange (RMCE). In other examples, suitable RRS can be selected from the group comprising FRT, F3, F5, FRT mutant-10, FRT mutant+10 and the mutants thereof, and in this scenario, the site-specific recombinase Flp recombinase or its derivative is used to

achieve RMCE. In yet another example, RRSs can be selected from the group comprising attB, attP and the mutants thereof, and in this case where the site-specific recombinase phiC31 integrase or its derivative is used to achieve RMCE.

[00129] In other embodiments, native cells are modified by a homologous recombination technique to integrate a nucleic acid sequence containing multiple exogenous nucleic acids into a specific site within an expression-enhancing locus.

[00130] For homologous recombination, homologous polynucleotide molecules (*i.e.* homologous arms) line up and exchange a stretch of their sequences. A transgene can be introduced during this exchange if the transgene is flanked by homologous genomic sequences. In one example, a recombinase recognition site can be introduced into the host cell genome at the integration sites via homologous recombination. In other examples, a nucleic acid sequence containing multiple exogenous nucleic acids, *e.g.*, multiple nucleic acids that together encode a bispecific antigen-binding protein, wherein the nucleic sequence is flanked by sequences homologous to the sequences at the target locus ("homologous arms"), is inserted into the host genome.

[00131] Homologous recombination in eukaryotic cells can be facilitated by introducing a break in the chromosomal DNA at the integration site. This may be accomplished by targeting certain nucleases to the specific site of integration. DNA-binding proteins that recognize DNA sequences at the target locus are known in the art. Gene targeting vectors are also employed to facilitate homologous recombination.

[00132] Gene targeting vector construction and nuclease selection to achieve homologous recombination are within the skill of the artisan to whom this invention pertains. In some examples, zinc finger nucleases (ZFNs), which have a modular structure and contain individual zinc finger domains, recognize a particular 3-nucleotide sequence in the target sequence (*e.g.* site of targeted integration). Some embodiments can utilize ZFNs with a combination of individual zinc finger domains targeting multiple target sequences.

Transcription activator-like (TAL) effector nucleases (TALENs) may also be employed for site-specific genome editing. TAL effector protein DNA-binding domain is typically utilized in combination with a non-specific cleavage domain of a restriction nuclease, such as FokI. In some embodiments, a fusion protein comprising a TAL effector protein DNA-binding

domain and a restriction nuclease cleavage domain is employed to recognize and cleave DNA at a target sequence within the locus of the invention (Boch J et al., 2009 *Science* 326:1509-1512). RNA-guided endonucleases (RGENs) are programmable genome engineering tools that were developed from bacterial adaptive immune machinery. In this system—the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) immune response—the protein Cas9 forms a sequence-specific endonuclease when complexed with two RNAs, one of which guides target selection. RGENs consist of components (Cas9 and tracrRNA) and a target-specific CRISPR RNA (crRNA). Both the efficiency of DNA target cleavage and the location of the cleavage sites vary based on the position of a protospacer adjacent motif (PAM), an additional requirement for target recognition (Chen, H. et al, *J. Biol. Chem.* published online March 14, 2014, as Manuscript M113.539726). Sequences unique for a specific targeting locus (such as SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3) can be identified by aligning many of these sequences to the CHO genome which can reveal potential off-target sites with 16-17 base pair match.

[00133] In some embodiments, a targeting vector carrying a nucleic acid of interest (e.g., a nucleic acid containing one or more RRSs optionally flanking one or more selectable marker gene, or a nucleic acid containing multiple exogenous nucleic acids which together encode a bispecific antigen-binding protein), flanked by 5' and 3' homology arms, is introduced into a cell with one or more additional vectors or mRNA. In one embodiment, the one or more additional vectors or mRNA contain a nucleotide sequence encoding a site-specific nuclease, including but not limited to a zinc finger nuclease (ZFN), a ZFN dimer, a transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, and an RNA-guided DNA endonuclease. In certain embodiments, the one or more vectors or mRNA include a first vector comprising a guide RNA, a tracrRNA and a nucleotide sequence encoding a Cas enzyme, and a second vector comprising a donor (exogenous) nucleotide sequence. Such donor sequence contains a nucleotide sequence coding for the gene of interest, or the recognition sequence, or the gene cassette comprising any one of these exogenous elements intended for targeted insertion. Where mRNA is used, the mRNA can be transfected into the cell by means of common transfection methods known to the skilled person and may encode an enzyme, for example a transposase or endonuclease. Although an

mRNA introduced into the cells may be transient and does not integrate into the genome, the mRNA may carry an exogenous nucleic acid necessary or beneficial for the integration to take place. In some instances, mRNA is chosen in order to eliminate any risk of long-lasting side effects of an accessory polynucleotide, where only short-term expression is required to achieve the desired integration of a nucleic acid.

Vectors For Site Specific Integration

[00134] Nucleic acid vectors are provided herein for introducing exogenous nucleic acids into an expression enhancing locus via site-specific integration. Suitable vectors include vectors designed to contain an exogenous nucleic sequence flanked by RRSs for integration via RMCE, and vectors designed to contain an exogenous nucleic sequence of interest flanked by homology arms for integration via homologous recombination.

[00135] In various embodiments, vectors are provided to achieve site-specific integration via RMCE. In some embodiments, vectors are designed to achieve simultaneous integration of multiple nucleic acids into a target locus. In contrast to sequential integration, simultaneous integration permits efficiency and rapid isolation of desirable clones that produce a bispecific antigen-binding protein.

[00136] In some embodiments, a set of vector is provided and includes two or more vectors, each vector containing at least two RRSs flanking one or more nucleic acids, wherein the nucleic acids in the vectors of the set together encoding a bispecific antigen binding protein.

[00137] In one embodiment, a set of vectors include a first vector comprising from 5' to 3': a first RRS, a first nucleic acid comprising a nucleotide sequence encoding a first LCF, and a third RRS; a second vector comprising from 5' to 3', the third RRS, a second nucleic acid comprising a nucleotide sequence encoding a first HCF, a second RRS; wherein either the first or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF; and wherein the first and second HCFs, and the first LCF, encode regions (e.g., variable regions) of a bispecific antigen-binding protein. In some embodiments, the nucleotide sequence encoding the second HCF is included in the first nucleic acid on the first vector (i.e., first LCF and second HCF on one vector), optionally placed, e.g., downstream of the nucleotide sequence encoding the first LCF; and in other embodiments, the nucleotide

sequence encoding the second HCF is included in the second nucleic acid on the second vector (first HCF and second HCF on one vector).

[00138] The nucleotide sequences encoding HCF can encode amino acids, e.g., amino acids or domain(s) from a constant region, or encode an entire constant region. In specific embodiments, a nucleotide sequences encoding an HCF or LCF can encode one or more constant domains, such as CL, CH1, hinge CH2, CH3, or combinations thereof. In certain embodiments, a nucleotide sequence encoding a HCF can encode a CH3 domain. For example, the nucleotide sequence encoding the first HCF can encode a first CH3 domain, and the nucleotide sequence encoding the second HCF can encode a second CH3 domain. The first and second CH3 domains can be the same, or differ in at least one amino acid. The differences in the CH3 domains or in the constant regions can take any of the formats for bispecific antigen-binding proteins described herein, e.g., differences that result in different Protein A binding characteristics, electrostatic steering, or in a "knob-and-hole" format. Independent of any amino acid sequence differences, the two HCF-coding nucleotide sequences can also differ in that one of the two nucleotide sequences has been codon modified.

[00139] In some embodiments, each HCF-coding nucleotide sequence is independently and operably linked to a transcriptional regulatory sequence including e.g., a promoter. In some embodiments, the promoters directing transcription of the two HCF-containing polypeptides are the same. In some embodiments, the promoters directing transcription of the two HCF-containing polypeptides, as well as the promoter directing transcription of the LCF-containing polypeptide, are all the same (e.g., a CMV promoter). In some embodiments, each HCF- or LCF-coding nucleotide sequence is independently and operably linked to an inducible or repressible promoter. Inducible or repressible promoters allow production to occur, for example, only in production phase (fed-batch culture) and not during growth phase (seed train culture). Inducible or repressible promoters also allow for differential expression of one or more genes of interest. In some embodiments, each HCF- and/or LCF-coding nucleotide sequence is independently and operably linked to a promoter upstream of at least one TetR operator (TetO) or Arc operator (ArcO). In still other embodiments, each HCF- and/or LCF-coding nucleotide sequence is independently and operably linked to a CMV/TetO or

CMV/ArcO hybrid promoter. Examples of hybrid promoters (also referred to as regulatory fusion proteins) may be found in International Publication No. WO03101189A1, published December 11, 2003 (herein incorporated by reference).

[00140] In some embodiments, the first nucleic acid in the first vector further comprises a 5' portion of a selectable marker gene, located at 5' to the third RRS in the first vector; and the second nucleic acid further comprises the remaining 3' portion of the selectable marker gene, located 3' to the third RRS in the second vector. In these embodiments, the first, second and third RRSs mediate site-specific integration of the first and second nucleic acids, which results in the joining of the 5' portion and the 3' portion of the selectable marker gene in properly and simultaneously integrated clones for convenient selection. In certain embodiments, the third RRS in the first vector is designed to be within a 5' portion of an intron of the selectable marker gene; and the third RRS in the second vector is designed to be within a 3' portion of an intron of the selectable marker gene. In still other embodiments, the third RRS in the first vector is designed to be between a promoter and the selectable marker gene to which it is operably linked (but it is separated from on the other vector); the third RRS in the first vector is designed to be 3' of a promoter; and the third RRS in the second vector is designed to be 5' of the selectable marker gene.

[00141] The set of vectors described above can include more than two vectors. For example, in addition to the two vectors described above, the set can include a third vector comprising at least two RRSs flanking a nucleotide sequence encoding a second LCF. The set of can also include a vector encoding one or more recombinases that recognizes the RRSs.

[00142] In other embodiments, vectors are provided to achieve site-specific integration via homologous recombination. In some examples, the polynucleotide sequence to be integrated into a host genome can be a DNA sequence, such as a RRS, or multiple RRSs flanking one or more selectable marker genes, for generating cells having one or more RRSs integrated in a desired locus for subsequent integration of nucleic acids encoding a bispecific antigen-binding protein. In other examples, the polynucleotide sequence to be integrated into the host genome includes multiple nucleic acids that together encode a bispecific antigen-binding protein. For example, the polynucleotide sequence includes nucleic acids encoding two different heavy chains and the common light chain of a bispecific antibody. In some embodiments, the

multiple nucleic acids that together encode a bispecific antigen-binding protein are each independently (i.e., separately) operably linked to regulatory sequences (such as a promoter, enhancer, a transcriptional termination sequence, or a combination thereof) - that is, the regulatory sequences (such as promoters) for each of the multiple nucleic acids are separate, which can be the same or different (i.e., containing the same or different nucleotide sequences). In instances where a nucleic acid among the multiple nucleic acids include multiple coding sequences, each coding sequence or each nucleotide sequence coding for the N-terminal portion of a polypeptide, is independently and operably linked to their own regulatory sequences (such as promoter).

[00143] It is well within the skill of the artisan to select sequences homologous to sequences within an expression enhancing locus and include the selected sequences as homology arms in a targeting vector. In some embodiments, the vector or construct comprises a first homologous arm and a second homologous arm, wherein the first and second homologous arms combined comprise a targeted sequence which replaces an endogenous sequence within the locus. In other embodiments, the first and second homologous arms comprise a targeted sequence which integrates or inserts within an endogenous sequence within the locus. In some embodiments, the homology arms contain a nucleotide sequence homologous to a nucleotide sequence present in SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3. In specific embodiments, the vector contains a 5' homology arm having the nucleotide sequence corresponding to nucleotides 1001-2001 of SEQ ID NO: 3, and a 3' homology arm having the nucleotide corresponding to nucleotides 2022-3022 of SEQ ID NO: 3. Homologous arms, for example a first homologous arm (also called 5' homology arm) and a second homologous arm (also called 3' homology arm) are homologous to a targeted sequence within the locus. The homologous arms from 5' to 3' may expand a region or targeted sequence within the locus that comprises at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In other embodiments, the total number of nucleotides of a targeted sequence selected for a first and second homologous arm comprises at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In some instances, the distance between the 5' homology arm and the 3' homology arm (homologous to the targeted sequence) comprises at least 5 bp, 10 bp, 20 bp,

30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, or at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In instances where nucleotides 1001-2001 and 2022-3022 of SEQ ID NO: 3 are chosen as 5' and 3' homology arms, the distance between the two homology arms can be 20 nucleotides (corresponding to nucleotides 2002-2021 of SEQ ID NO: 3); and such homology arms can mediate integration of an exogenous nucleic acid sequence within a locus comprising SEQ ID NO: 3, e.g., within nucleotides 1990-2021 or 2002-2021 of SEQ ID NO: 3, and a simultaneous deletion of nucleotides 2002-2021 of SEQ ID NO: 3.

[00144] The vectors disclosed herein for introducing exogenous nucleic acids for site-specific integration into an expression enhancing locus can include additional genes and sequences for directing the expression of exogenous nucleic acids of interest and encoded polypeptides and for the selection and identification of cells into which the exogenous nucleic acids of interest have successfully integrated. Such additional sequences include, for example, transcriptional and translational regulatory sequences, selectable marker genes, and the like, also described hereinbelow.

Regulatory Sequences

[00145] The vectors disclosed herein for introducing exogenous nucleic acids into an expression enhancing locus in a site-specific manner, and the cells obtained as a result of site-specific integration, can include regulator sequences for directing the expression of exogenous nucleic acids of interest and encoded polypeptides. Regulatory sequences include transcriptional promoters, enhancers, sequences encoding suitable mRNA ribosomal binding sites, and sequences that control the termination of transcription and translation. Transcriptional and translational control sequences may be provided by viral sources. For example, commonly used promoters and enhancers are derived from viruses such as polyoma, adenovirus 2, simian virus 40 (SV40), mouse or human cytomegalovirus (CMV), CMV immediate early (CMV-IE) or CMV major IE (CMV-MIE) promoter, as well as RSV, SV40 late promoter, SL3-3, MMTV, ubiquitin (Ubi), ubiquitin C (UbC), and HIV LTR promoters. Viral genomic promoters, control and/or signal sequences may be utilized to drive expression, provided such control sequences are compatible with the host cell chosen. Non-viral cellular

promoters can also be used (e.g., the β -globin and the EF-1 α promoters), depending on the cell type in which the proteins of interest are to be expressed. DNA sequences derived from the SV40 viral genome, for example, the SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide other genetic elements useful for expression of a exogenous DNA sequence. Early and late promoters are particularly useful because both are obtained easily from the SV40 virus as a fragment that also comprises the SV40 viral origin of replication (Fiers *et al.*, *Nature* 273:113, 1978). Smaller or larger SV40 fragments may also be used. Typically, the approximately 250 bp sequence extending from the Hind III site toward the BglII site located in the SV40 origin of replication is included. Inducible promoters (induced by a chemical compound, cofactor, regulatory protein, for example) can be used and are particularly useful for allowing the production of antigen binding proteins to occur only in production phase (fed-batch culture) and not during growth phase (seed train culture). Examples of inducible or repressible promoters include alcohol dehydrogenase I gene promoters, tetracycline-responsive promoter systems, glucocorticoid receptor promoters, estrogen receptor promoter, ecdysone receptor promoters, metallothionein-based promoters, and T7-polymerase based promoters. Sequences suitable for the expression of multiple transcripts via a bicistronic vector have been described previously (Kim S. K. and Wold B. J., *Cell* 42:129, 1985) and can be used this invention. Examples of suitable strategies for multicistronic expression of proteins include the use of a 2A peptide (Szymczak *et al.*, *Expert Opin Biol Ther* 5: 627–638 (2005)) and the use of an internal ribosome entry site ("IRES"), both well known in the art. Other types of expression vectors will also be useful, for example, those described in U.S. Pat. No. 4,634,665 (Axel *et al.*) and U.S. Pat. No. 4,656,134 (Ringold *et al.*).

Selectable markers

[00146] The vectors disclosed herein for introducing exogenous nucleic acids into an expression enhancing locus in a site-specific manner, and the cells obtained as a result of site-specific integration, can include one or more selectable markers genes.

[00147] In some embodiments, a selectable marker gene confers drug resistance, such as, for example, those described in Table 1 of Kaufman, R. J. (1988) *Meth. Enzymology* 185:537, and include DHFR-MTX resistance, P-glycoprotein and multiple drug resistance (MDR)-

various lipophilic cytotoxic agents (e.g., adriamycin, colchicine, vincristine), and adenosine deaminase (ADA)-Xyl-A or adenosine and 2'-deoxycoformycin. Other dominant selectable markers include microbially derived antibiotic resistance genes, for example neomycin, kanamycin or hygromycin resistance. Several suitable selection systems exist for mammalian hosts (Sambrook *supra*, pgs 16.9-16.15). Co-transfection protocols employing two dominant selectable markers have also been described (Okayama and Berg, *Mol. Cell Biol* 5:1136, 1985).

[00148] In other embodiments, a selectable marker gene encodes a polypeptide that provides or is capable of generating a detectable signal for the recognition of gene cassettes that have or have not been successfully inserted and/or replaced, as the case may be. Suitable examples include a fluorescent marker or protein, an enzyme that catalyzes a chemical reaction that generates a detectable signal, among others. Examples of fluorescent markers are well-known in the art, including, but not limited to *Discosoma* coral (DsRed), green fluorescent protein (GFP), enhanced green fluorescent protein (eGFP), cyano fluorescent protein (CFP), enhanced cyano fluorescent protein (eCFP), yellow fluorescent protein (YFP), enhanced yellow fluorescent protein (eYFP) and far-red fluorescent protein (e.g. mKate, mKate2, mPlum, mRaspberry or E2-crimson. See also, e.g., Nagai, T., et al. 2002 *Nature Biotechnology* 20:87-90; Heim, R. et al. 23 February 1995 *Nature* 373:663-664; and Strack, R.L. et al. 2009 *Biochemistry* 48:8279-81.

Systems for Making Bispecific Antigen-Binding Proteins

[00149] In a further aspect, this disclosure provides systems that include a combination of a cell and one or more vectors, and that can be utilized to make cells having integrated within an expression enhancing locus exogenous nucleic acids that together encode a bispecific antigen binding protein. The systems can be provided in the form of a kit, for example.

[00150] In some embodiments, a system is designed to permit efficient vector construction and simultaneous integration of multiple exogenous nucleic acids via RMCE into a specific site within an enhanced expression locus. Simultaneous integration permits rapid isolation of desirable clones, and the use of one enhanced expression locus is also important for creation of a stable cell line.

[00151] In some embodiments, a system is provided that includes any one of the set of vectors described above designed to integrate multiple exogenous nucleic acids via RMCE, and a cell containing RRSs that are integrated at a specific site within an enhanced expression locus and that match the RRS in the set of vectors. For example, a system includes a cell and a set for vectors, wherein the cell contains, integrated within an enhanced expression locus of its genome from 5' to 3': a first RRS, a first exogenous nucleic acid, a second RRS, a second exogenous nucleic acid, and a third RRS, wherein the three RRSs are different from one another; wherein the set of vectors includes a first vector comprising from 5' to 3', the first RRS, a first nucleic acid comprising a nucleotide sequence encoding a first LCF (e.g., a first VL), and the second RRS; a second vector comprising the second RRS, a second nucleic acid comprising a nucleotide sequence encoding a first HCF (e.g., a first VH), and the third RRS; and wherein either the first nucleic acid or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF (e.g., a second VH). Upon introduction of the vectors into the cell, the first and second nucleic acids in the vectors integrate into the enhanced expression locus through recombination mediated by the first, second and third RRSs. To facilitate screening of transfecants having nucleic acids properly integrated from the vectors into the locus, the first exogenous nucleic acid in the cell of the system can include a first selectable marker gene, and the second exogenous nucleic acid in the cell can include a second selectable marker gene, wherein the first and second selectable marker genes are different from each other, and are also different from any selectable marker gene provided by the vectors; and in specific embodiments, the first and second selectable marker genes encode fluorescent proteins (which can provide negative selection), and the first and second nucleic acids on the vectors provide an additional selectable marker gene in a split format to provide positive selection. Negative selection alone can provide rapid clone isolation, although the efficiency of isolating clones with intended recombination may be limited (about 1%). Negative selection coupled with positive selection (a new fluorescence, or a resistance to a drug or antibiotic) can significantly improve the efficiency of isolation of positive clones (to about 80%).

[00152] The systems can include additional components, reagents, or information, for examples, protocols for introducing the vector(s) in a system into the cell of the system by

transfection. Non-limiting transfection methods include chemical-based transfection methods include the use of liposomes; nanoparticles; calcium phosphate (Graham et al. (1973) *Virology* 52 (2): 456–67, Bacchetti et al. (1977) *Proc Natl Acad Sci USA* 74 (4): 1590–4 and, Kriegler, M (1991) Transfer and Expression: A Laboratory Manual. New York: W. H. Freeman and Company. pp. 96–97); dendrimers; or cationic polymers such as DEAE-dextran or polyethylenimine. Non chemical methods include electroporation; sono-poration; and optical transfection. Particle-based transfection include the use of a gene gun, magnet assisted transfection (Bertram, J. (2006) *Current Pharmaceutical Biotechnology* 7, 277–28). Viral methods can also be used for transfection. mRNA delivery includes methods using TransMessenger™ and TransIT® (Bire et al. *BMC Biotechnology* 2013, 13:75). One commonly used method of introducing heterologous DNA into a cell is calcium phosphate precipitation, for example, as described by Wigler et al. (*Proc. Natl. Acad. Sci. USA* 77:3567, 1980). Polyethylene-induced fusion of bacterial protoplasts with mammalian cells (Schaffner et al., (1980) *Proc. Natl. Acad. Sci. USA* 77:2163) is another useful method of introducing heterologous DNA. Electroporation can also be used to introduce DNA directly into the cytoplasm of a host cell, for example, as described by Potter et al. (*Proc. Natl. Acad. Sci. USA* 81:7161, 1988) or Shigekawa et al. (*BioTechniques* 6:742, 1988). Other reagents useful for introducing heterologous DNA into a mammalian cell have been described, such as Lipofectin™ Reagent and Lipofectamine™ Reagent (Gibco BRL, Gaithersburg, Md.). Both of these commercially available reagents are used to form lipid-nucleic acid complexes (or liposomes) which, when applied to cultured cells, facilitate uptake of the nucleic acid into the cells.

Methods for Making Bispecific Antigen-Binding Proteins

[00153] This disclosure also provides methods of making bispecific antigen-binding proteins. By utilizing the present methods, a bispecific antigen-binding protein (e.g., a bispecific antibody) can be produced at high titers and/or high specific productivity (pg/cell/day). In some embodiments, a bispecific antigen-binding protein is produced at a titer of at least 5 mg/L, 10 mg/L, 15 mg/L, 20 mg/L, 25 mg/L, 30 mg/L, 35 mg/L, 40 mg/L, 45 mg/L, 50 mg/L, or greater. In some embodiments, a bispecific antigen-binding protein is produced at a ratio of the bispecific antigen-binding protein titer versus the total antigen-binding protein titer of

at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, or higher. In some embodiments, a bispecific antigen-binding protein is produced at a specific productivity of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 picogram/cell/day, or higher, determined based on total antigen-binding proteins (in pg) produced per cell per day.

[00154] In one embodiment, the method utilizes a system disclosed herein and introduces the vectors in the system into the cell of the system by transfection. Transfected cells where the exogenous nucleic acids have been properly integrated into a target enhanced expression locus of the cell through RMCE can be screened and identified. The two HCF-containing polypeptides and at least one LCF-containing polypeptide can be expressed from the integrated nucleic acids, and the bispecific antigen-binding protein containing all three polypeptides can be obtained from the identified transfected cell, and purified using known methods.

[00155] In some embodiments, a method includes (i) providing a system that includes a cell and a set for vectors, wherein the cell contains, integrated within an enhanced expression locus of its genome from 5' to 3': a first RRS, a first exogenous nucleic acid, a second RRS, a second exogenous nucleic acid, and a third RRS, wherein the three RRSs are different from one another; wherein the set of vectors includes a first vector comprising from 5' to 3', the first RRS, a first nucleic acid comprising a nucleotide sequence encoding a first LCF (e.g., a first VL), and the second RRS; a second vector comprising the second RRS, a second nucleic acid comprising a nucleotide sequence encoding a first HCF (e.g., a first VH), and the third RRS; and wherein either the first nucleic acid or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF (e.g., a second VH); (ii) introducing the vectors simultaneously into the cell; and (iii) screening for transformed cells in which the first and second nucleic acids in the vectors have simultaneously integrated into the enhanced expression locus through recombination mediated by the first, second and third RRSs.

[00156] In a specific embodiment of the method, to facilitate screening of transformants having nucleic acids properly integrated from the vectors into the locus, the first exogenous nucleic acid in the cell of the system can include a first selectable marker gene, and the second exogenous nucleic acid in the cell can include a second selectable marker gene, wherein the first and second selectable marker genes are different from each other; and the

first and second nucleic acids on the vectors together encode an additional selectable marker in a split format where a complete sequence encoding the additional selectable marker gene is provided following simultaneous integration. Screening for transformants can be conducted to select against the first and second selectable markers (negative selection), and for the additional selectable marker (positive selection).

[00157] In another embodiment, the method simply utilizes a cell having an exogenous nucleic acid sequence integrated at a specific site within an enhanced expression locus of the cell, wherein the exogenous nucleic acid sequence encodes a bispecific antigen-binding protein, and expresses the bispecific antigen-binding protein from the cell. The cloned expression cassette is contiguous within the specific integration site.

[00158] The present description is further illustrated by the following examples, which should not be construed as limiting in any way. The contents of all cited references (including literature references, issued patents, and published patent applications as cited throughout this application) are hereby expressly incorporated by reference.

EXAMPLES

Example 1: Cloning of bispecific antibody expression plasmids:

[00159] The heavy and light chain components of bispecific antibodies may be cloned from hybridoma cells, B cells, plasma cells, or recombinant antibody gene libraries using methods known in the art. For example, antibodies may be cloned from hybridoma or B cells by five prime RACE PCR, or PCR using primers against leader peptides, framework 1 sequences, framework 4 sequences, or constant region sequences. Alternatively, antibody genes or mRNAs in antibody-expressing cells may be sequenced by next-generation sequencing and subsequently identified through bioinformatics. It is also feasible to sequence antibody proteins and clone the corresponding antibody genes by synthetic DNA technology. Recombinant antibody libraries, such as yeast or phage libraries, are also sources of antibody genes.

[00160] The CHO expression cell lines RSX4189-1, RSX4187-1, RSX4191-1, and RSX4188-1 each produce a bispecific antibody comprised of three distinct polypeptides:

AbC1, AbC2, and AbC3 (Figure 5). To generate plasmids for construction of RSX4189-1, the AbC1 plasmid was linearized by digestion with Mfe I, which was 3' to the AbC1 gene. The AbC2 expression cassette excised from the AbC2 plasmid by Mfe I digestion was ligated to the Mfe I site of the linearized AbC1 plasmid. The ligation products were transformed into DH10B *E. coli*. After transformation and growth in Ampicillin-containing LB media, individual *E. coli* colonies were analyzed for harboring the desired plasmid containing AbC1 and AbC2 genes. Sequences of maxi-prep DNA for the AbC3 plasmid and the AbC1-AbC2 dual expression plasmid were confirmed by Sanger sequencing. These two plasmids, together with the Cre expression plasmid pRG858, were transfected into EESYR® host cells harboring RRS1 and RRS3 sites at the EESYR® locus using lipofectamine. The transfected cells were selected with antibiotics for 12 days, and recombinant cells were subsequently pooled as RSX4189-1.

[00161] To generate plasmids for construction of RSX4187-1, the AbC3 expression cassette flanked by Mlu I and Nhe I sites was cloned into the Mlu I and Spe I sites in the AbC2 plasmid, 3' to the AbC2 gene. The combined AbC2-AbC3 plasmid, the AbC1 plasmid, and the Cre plasmid pRG858 was co-transfected into EESYR® host cells using lipofectamine. The cells that had undergone RMCE were pooled as RSX4187-1.

[00162] To generate plasmids for construction of RSX4191-1, the AbC3 expression cassette was cloned into the Mfe I site in the AbC1 plasmid, 3' to the AbC1 gene. The combined AbC1-AbC3 plasmid, the AbC2 plasmid, and the Cre plasmid pRG858 was co-transfected into EESYR® host cells using lipofectamine. The cells that had undergone RMCE were pooled as RSX4191-1.

[00163] To generate plasmids for construction of RSX4188-1, the AbC2 expression cassette flanked by Mlu I and Nhe I sites was cloned into the Mlu I and Spe I sites in the AbC3 plasmid, 3' to the AbC3 gene. The combined AbC3-AbC2 plasmid, the AbC1 plasmid, and the Cre plasmid pRG858 was co-transfected into EESYR® host cells using lipofectamine. The cells that had undergone RMCE were pooled as RSX4188-1.

Example 2: Expression of bispecific antibody from the EESYR® locus

[00164] The bispecific antibody expression cell lines RSX4189-1, RSX4187-1, RSX4191-1, and RSX4188-1 were cultured in suspension in serum-free medium. To quantitate their bispecific antibody expression levels, the cell numbers of the cultures were counted on a Guava flow cytometer and fresh shaker flask cultures containing 2 million cells per ml medium were started. 4 days later, the spent media were harvested after centrifugation to remove the cells. The bispecific antibody titers were determined using a protein A HPLC assay that was specific to the bispecific antibody. The titers of the bispecific antibody protein expressed from RSX4189-1, RSX4187-1, RSX4191-1, and RSX4188-1 were 37.8 mg/L, 40.5 mg/L, 48.3 mg/L, and 21.8 mg/L, respectively. The total titer of all antibody proteins (including the bispecific antibody protein and the monospecific antibody proteins) expressed from these cell lines, and the ratio of the bispecific antibody protein titer versus the total antibody protein titer, are shown in the table below.

AbPID/RSX	Bispecific Titer (mg/L)	Total Titer (mg/L)	% Bispecific
RSX4187-1	40.5	211.7	19.10%
RSX4188-1	21.8	218.3	10.00%
RSX4189-1	37.8	199.5	19.00%
RSX4191-1	48.3	172.9	27.90%

SEQUENCE LISTING

SEQ ID NO: 1

13515 bases

DNA

Cricetulus griseus

tctagaaaaca	aaacccaaaaaa	tattaagtca	ggcttggctt	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaaattc	ctggcttct	aaggctttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaatctg	aaacataaggt	gatgtatctg	ccataactgca	tgggtgtgt	180
tgtgtgtgt	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaaaaca	240
cttcaaagac	aggagagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtgag	gactgaacca	ttaaacacaa	gtggtaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggtca	ggagtggttc	tctggaggct	420
tctgccctg	gaggttccctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgacta	gagagagctg	agggaccact	ggactgaggc	ctagatgctc	540
agtcaagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcaggtgt	600
gaggggtcag	gggcaagtt	gtgggagagg	tcttccagat	gaagtagcag	gaacggagac	660
gcactggatg	gccccacttg	tcaaccagca	aaagcttga	tcttgttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggtttttaaa	aggctttgtg	ttacctaatac	acttcttatta	780
gtcagatact	ttgttaacaca	aatagtagact	tggcctgtat	tttagaaaact	tctgggatcc	840
tgaaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccaagcca	ggccctggac	900
ccgggtccat	tcatgcaaat	actcagggac	agattctca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcattttac	tacaagtca	catgagtcag	gaggtgtgt	1020
ttgcacagtg	tgactaagtg	atggagtgtt	gactgcagcc	attcccccgc	ccagcttgtg	1080
agagagatcc	ttttaaattg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacaca	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagccttc	cttacctact	accttccagt	gatcagggtt	1260
gactgtcctg	tgtgtactg	gaccctgaat	gtccccaccc	ctgtccctgt	tctttacga	1320
ttctgacatt	ttaataaaat	tcagcggctt	cccctctgtct	ctgtgcctag	ctatacctt	1380
gtactctgca	ttttggttt	tgtgacattt	ctctgtact	ctgctacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatccccca	aactgttgc	cagtgtttgg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtqgt	ccaggctgt	1620
gacttcactc	cccccaagtga	gggaggccat	ggtgagtaca	cacacccac	acatactata	1680
tccaaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacat	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatac	cacacataca	cacccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacat	acataggcac	acattcacac	accacacata	tacatttgc	tatgcataca	1920
tgcatacaca	cacaggcaca	cacacaccac	acacatgcac	tgtgtacgca	cacatgcata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgttatatgca	cacatgcac	cacacataca	taggcacaca	tagagcacac	2100
acatacattt	gtgttatgcac	acatgcacac	accaatcaca	tggaaagact	caggttcttc	2160
actaagggttc	acatgaactt	agcagtctt	ggttatctcg	tgaaacttgg	aatattgtct	2220
tggagaagag	gaagcgttgg	tttgagccct	ggcagcaatt	aacccccc	agaagaagta	2280
gttttaaaaa	tgagagggtc	tcaatgtgga	accgcagg	cgccagttca	gagaagagac	2340
ctacccaagc	caactgagag	caaaggcaga	ggatgaacc	tggatgttag	tttgaaccc	2400
tgttaccagct	gggcttcatg	ctatttgtt	atatcttata	taaatattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaaatgt	atgggcactg	tatgtgttct	tgggtccgg	2520
ggaggccagg	agagggcatg	gatcctccgg	agctggcggt	tgagacagtt	gtgacccaca	2580
gtgtgggtc	tggaaactgg	gtcttagtgt	tccgcaagtg	cagctggggc	tcttaacctc	2640
tgagccatcc	ctccagctc	aagaaactta	ttttcttagg	acatggggga	aggatccag	2700
ggctttaggc	ttgtttgttc	agcaaatact	ctttcgtgt	atttgaatt	ttatttatt	2760

ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcatcaaaat	2820
cctcctgtct	cagtctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
cttttagagtc	cccaagccctt	ctggacactt	gttccaagta	taatataatat	atataatatat	2940
atataatatat	atataatatat	atataattgtg	tgtgtgtgtt	tgtgtgtgt	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgcttttaat	ttatTTTTA	attaaaaatg	3060
attagctaca	tgtcacctgt	atgcgtctgt	atcatctata	tatccttcct	tccttcctc	3120
tctttctctc	ttctctctct	cacccccaag	cacatctatTTT	caaatttcct	tgccgaggag	3180
atgccaagag	tctcgTTggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aagggggaca	gactgggtgt	ggtctttgga	gagctcagga	gaatagcagc	gatcttcct	3300
gtcccctggg	tcaccttta	cagccaaacac	cattttgtgg	cctggcagaa	gagttgtcaa	3360
gctggtcgca	ggtctgccac	acaacccaa	tctggcccca	agaaaaggca	cctgtgtgt	3420
actctgggg	taaaggcgct	gcctggctgt	ctccagctgg	acttggaaact	cccgTTtaat	3480
aaagagtct	gcaaaataat	acccgcagag	tcacagtgc	aggTTcccgt	gttttcctga	3540
agcgccagc	acgggttccc	taggaaatgg	ggccttgctt	gccaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcaact	tgcgttgcca	ctggatgaa	atggaaaaaaa	3660
aaaaaaagaag	aagtgtctct	ggaagcgggc	gcgcacac	aaacccgcaa	cgattgtgt	3720
aacactctcc	attgagaatc	tggagtgcgg	ttggcctcta	ctggggagct	gaagacagct	3780
agtgggggcg	gggggaggac	cgtgctagca	tccttccacg	gtgctcgctg	gctgtgtgtc	3840
atgcccggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttgggtggaa	ctgacaatca	3900
acgaaatcac	ttcgattgtt	ttcctcttt	tactggatt	cttggatttg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
gggggaggagg	ggaggaggggg	aagggatgga	ggaaaaatact	aacttttcta	attcaacatg	4080
acaaaagatc	ggagaaagt	caccgctagt	gaccggagg	aggaatgccc	tattgggcat	4140
tatattccct	gtcgcttaat	ggaatcaaac	tcttggttcc	agcaccagg	attctgagcc	4200
tatcctattc	aagacagtaa	ctacagccca	cacggaaagag	gctataacaac	tgaagaaata	4260
aaattttcac	tttatttcat	ttctgtact	gcatgttac	atgtagagag	ccacctgtgt	4320
ctagggcgt	atgtgtgtgg	cagtagagtt	ctgagccgt	taacttggaa	aacccagaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttatTTtagag	acagggtctc	atactgttagt	ccaaagctago	ctccagctca	cagaaattct	4500
cctgttccgg	tttccaaagt	actggagtt	tgagtgtgt	ttaatttgaac	gctaagaatt	4560
tgctgattga	agaaaaacctc	aagtgggtt	ggctaattccc	cacgacccca	gaggctgagg	4620
caggaggaat	gagagaattc	aaggTTTgCC	agagccacag	ggtgagctca	atgtggagac	4680
tgtgagggtg	agctcaatgt	ggagactgtg	agggtgagct	caatgtggag	actgtgaggg	4740
tgagctcaat	gtggagactg	tgagggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	agggtgtgtt	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccggtcc	atctgtggcc	4920
ctgcaacagg	aagggaggga	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagataggg	aatgagagag	gaaggaagaa	acgggaagaa	attcagactc	cttcctgagt	5040
tccgccaacg	cctagtgaca	tcctgtgcac	accctaaggt	ggcctttgtg	tggactggc	5100
ttgggtggtc	gggaaaggca	tttcagctt	gttgcagaac	tgccacagta	gcatgtgtgg	5160
tccgtaaaag	tttctgccc	ttaacaagaa	gtctctacta	cttgtgacct	caccagtgaa	5220
aatttcttta	attgtctct	ggtgttctgg	tttttgcatt	tttgcatttcta	aggatacatt	5280
cctgggtgtat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttggat	ttggaaagat	5340
gatttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttggaa	aatgtcctcc	5400
cgcggcccc	aattttggct	tggcaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460
gaggaggcat	ggggtcctag	gaggacagga	agcagaagga	gagagctgg	ctgacagcct	5520
gcaggcatgt	cacagttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtggcgtct	gaggcaactt	gacctcagcc	agatggtatt	5640
tgaataaccc	gtcttttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgtatg	5700
aactctca	ttgcctgagt	catgtgcggc	catgcccagg	ggtcaggctg	acactcaact	5760
aaaaaaacaag	tgagaaattt	aagacaatcc	gtggtggcag	ctacttggaa	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaaagc	catttgcatt	aggcacagtt	atcttgaatg	5880
catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacattc	gagttgagtc	5940
attcagtgt	ggaaaccccg	acgcttccat	cccctaaaag	gaacatctt	ctctcagtca	6000
aaatggaaat	aaaaatttggg	gcttgaattt	ggcaaatgat	tcagaactct	gtgttaggtat	6060

tttcacacgc acagtggata	attttcatgt tggagtttat	ttgtgctaaa	aggcagaaaa	6120
gggtaaaaag cacatcttaa	gagttatgag gttctacgaa	taaaaataat	gttacttaca	6180
gttattccctt aatttagtacc	cccttccacc tgggttaatt	tcctgagata	gtcagtgggg	6240
aaaagatctc tccttcttct	ctttctcccc ctccccctct	ctccctccct	ccctccctcc	6300
ctccctccctc tccctccctc	cccctttctt	tcttcttgc	tccttctcc	6360
tctccctttc ttcttcattt	attctaaatgt	gctttaaaca	gcacaccaat	6420
taacgggaaa acacaggctc	aaggcgtta	gagaagattg	atctgtgttc	6480
aattcagagg tgggtgaaga	taaaaggcaa	acatttgagg	ccatttcctt	6540
gcacttagga agtggAACAT	gcctaacta	ctgggttgta	ccaccttcc	6600
ctgtttggga agctccctgg	caaccgattc	tggcatctca	ttggtcagag	6660
tggtaacttctt atttgcaaag	aaggctgtaa	ctttagctt	taaaagectc	6720
agaagggaga aaggatatgg	ctagacatat	ctaataagact	taaccactgt	6780
agtatgaatc agatagaacc	tatTTTAAAC	tcagtttga	aaaaaataat	6840
atTTTGTGTgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	6900
tagcaggtgc tggaggaggc	cagaagaggg	caccagatct	cctggaaactg	6960
tggttatgag ctgcctgatg	tgggtgcgtt	gaactgaact	ctcgtgttct	7020
caactgttct cttaactgtat	gagccatctc	tccagcccc	cccataattt	7080
ttttagtaaa ttttattcat	aatcaattat	cacagtataa	aacaatgtatt	7140
catatacata tcaaggatga	cagtggggg	gatatgtgt	tgtgtgtgt	7200
tgtgtgtgt	gtgtgtgtc	tttttaagaa	ggtgcctatag	7260
tctctgaagg atttcaaagg	aatgagacat	gtctgtctc	caggaaccct	7320
ttgggaatct gacccaaattg	aggtattctg	aggaactgaa	tgaagagctc	7380
gtcttaaacc caaatgtgct	gtcttagagaa	agtcaacgctc	atcagtgagc	7440
tttactgagc ggaagacaag	cgctcttga	ttaagtggc	tgcacagtc	7500
agtggagcgt gtgtcaqgt	ctgaggcagt	cttgcgttgc	cagctgtqat	7560
gaaagggtgg agatggaggc	agggtggag	caggcgtatg	gttcagacta	7620
gcacaccagc tgggtgactt	gtggctgtg	ggtcaggcgt	tgtaaacgc	7680
ggcagtcaca ttggttgaag	ctgaatgggt	gaggcaacac	agagagtgc	7740
agtaccacat cttccccgac	ccaggtcaact	tctgggttat	agctgagact	7800
tgcaaccagc tggtagagc	ttcaggaaa	acttgatgtc	tgcattgtc	7860
gattcggtagc atctggagaa	aatttataat	gctggctcag	tcaagcactg	7920
ccttggctt gggagctaca	tgacatttgc	ttttaggcag	acttttttt	7980
caattcccaag ataaccaata	tggaggctca	atattaatta	taaatgctcg	8040
caggcttgtt actagctaac	tcttccaact	taaatgaacc	catttctatt	8100
tgccacgtga cttaacccct	tacttccctgt	ttcctctctt	tgtctgactc	8160
cttcccagag tccttagtct	ggttctctg	cctaaccctt	tcctgcccag	8220
agcatttata attaatatta	agtctccag	ttagactctc	atccagggag	8280
ctccccccctc ctcattgcca	tccgtgttctt	cctttccct	cgcttcccc	8340
gctttccctc ctccacccct	ccttccatag	tattgtggc	aagggtgttc	8400
gagtgcctt aggcattgca	agaaaccagt	tagatgtc	tgtgaggggt	8460
agcgtatggac acaattcaag	ccacagagt	aagacggaag	gatgcactgt	8520
aacttctggg gcagaatcac	agggtgagg	tctgacttga	gggcgaagag	8580
agggagttag tttgtctgag	ctagaagcta	cggcccacct	tttggtagca	8640
caagcatgt ttgttaatca	tgtggatct	gatTTTCTC	taaatctatg	8700
aagaaaatgt gaattctcac	attaaaattt	agatatacgt	cttttgggtgg	8760
aaaaaatctc aagaatatgg	atttctgggg	gccggagaga	tggctcagag	8820
ctgggtgctc ttcttagacat	tctgagttca	attcccagca	accacatgg	8880
catctgtaat gcgacccgtt	gccatcttct	gacatgc	gatacatgca	8940
tgtatacata gtaaatttgc	aaatctttt	ttaaaaagag	tatggattct	9000
ggtggcgcac gccttaatc	ccagcactt	ggaggcagag	gcaggtggat	9060
tcgagaccag cctggcttat	aagagctagt	tccaggacag	cctccaaagc	9120
ccctgtctcg aaaaacccaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	9180
taagaaagcc gtaacagctg	gagctgtgt	cgagttcag	cgtggacta	9240
cattcatgt gaaacacccc	aggattttt	cttagtatct	agttccatt	9300
gaccggctct tatgtctcc	aggctggcct	caaactgctg	atctcccg	9360

caagtccctgg	gactacttgg	ctcataaaaac	agttttgtc	gggctccctg	aagttatgg	9420
tgtacaaaacc	gtgggggtca	atatactcac	ttggggcagag	agagaagg	tgaatcc	9480
acaatgactg	catctcagga	cagttggaa	gaggacaatg	gcagaagg	ttagaaa	9540
tagactggag	ggtggaaaag	cagcaggaac	agagaaaaca	aacagga	ttgctatcc	9600
ggccactct	ggagtccctgt	ggcaagatgg	aagcgggcta	gggaaata	tttgtgtac	9660
tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	caatgcctat	caatgtt	9720
ggggaaat	gtataaccaca	ttgattctgg	gagcaattct	cagtatctgg	cctagagaaa	9780
ggaatggccc	ctgcagaata	gacagagtga	atggtgcct	ttatcatttgc	ctaaagt	9840
ggagaaataa	acatccctcc	atagagtttgc	aggttaatga	accccac	tcatctgt	9900
cgtggtgag	gcctggccaa	cagttaaaaa	gattagacac	ggacaaagtc	tgaaggaaac	9960
acctcgaata	ggaagaggag	agccaccc	ttctgttaact	ttcctcaagg	ggaagatgtt	10020
ccaagagtgg	gaataaaatgg	tcaaaggggg	gatttttaat	tagaaaacg	atttcctgt	10080
tcacttgc	aactggaggt	tgatttgggg	cataggacaa	tagattgtat	gttttgcaaa	10140
aagctgttc	aaagcagaga	aatggatag	agacaattat	gtagcgagga	gggagggtgg	10200
ggcgaagatg	gagacagaga	agtggaa	gactttagg	aagaggaaca	tagaccac	10260
gggcggggcg	gggggcaggg	gcggggggcg	gggctcaaag	gagggcagtgg	gaacgttgc	10320
agtgttcgca	gcgttaagcgt	gaatgtgca	gcgtcttgc	ggtgtgtgc	caggagtagc	10380
gtggctggct	tgtgtgtgc	ttgtatccc	agtcttgc	gtttccacac	tgttccac	10440
tgggtgtgat	tttcctcgg	agagcatgag	ggctctgc	tcccccac	ctccccagcg	10500
ttcgttggta	tttgtttcca	agatgttagt	gggtgagaca	aagcctct	gttatttgc	10560
ctttaacagg	tgacaaaaaa	agctcaacca	ggagacattt	ttgccttctt	ggaaggtaat	10620
gtcccatgt	agagcaatgg	gaccatctc	taaggtgagg	ctactcttgc	agtttgcacc	10680
cagctttct	gatgcaggaa	ggaagttgt	gggcaagcaa	gactgttgc	ttcttgc	10740
ggacacatcc	tgcacacaaa	ggctcaggag	gggagaaggc	tgtttgatgt	ttagcact	10800
ggaaggcccc	tgatgcac	gtgattag	gtctccatct	gtggagcaga	cacqgactaa	10860
ctaaaaacca	gtgttttaa	attgtcaagc	ctttaaggt	aggaaattga	cttatttgc	10920
tgggcatac	gtagagcaag	tgctctgc	tggccaacc	cccggctct	gttcttaggc	10980
accagaatgg	cctagaacta	actcacaatc	ctcccaattcc	aggtctcagg	tgctagaat	11040
aaccactata	ccagccgtcc	tgcctgc	cctgccttcc	taaattttaa	atcatggg	11100
gtaggggaga	atacacttat	cttagttag	gtttctatttgc	ctgtgaagag	acaccatgag	11160
catggcaact	cttataaagg	aaaacattt	gttgggtggc	agtttcagag	gttttagtac	11220
attgtcatca	tggctggaa	catgtggca	tgcagacaga	catggtgct	gagaaaggga	11280
tgagagtct	acatcttgc	ggcaacagga	cctcagctga	gacactggct	ggtaccctga	11340
gcataggaaa	cctcacagcc	caccctcaca	gtgacatatt	tccttcaaca	aagccataacc	11400
tcctataagt	gccactccct	atgagatgc	agggcaattt	acattcaa	tgcataaaca	11460
ctttaaagta	ttttat	attattgtaa	attatgtat	tagctgggt	gtggcagcc	11520
agggtcacgc	ctttaatccc	agcacttgg	aggcagaggc	agatggatct	ctgtgagtt	11580
aagaccagcc	tggctataa	gagctagtt	caaggaagga	tatacaaaga	acagttctag	11640
gatagcctc	aaagccacag	agaagtgc	tcttggaaac	caaaaattgt	gtgggac	11700
gtctctgc	tggctgc	ccactcccc	agagctggac	tcttggtcaa	cactgaat	11760
gctgaaaaat	aaactccctgg	attcctct	tgtacatgg	gcccgaa	aggcgcc	11820
ttgtcttc	gcaggattgc	catagactt	ttctgtgt	ccaccat	agactga	11880
agagatggca	gtggcagaga	ctgggaaggc	tgcacacaaa	acaggaagtt	attgcacc	11940
ggaatagtc	tggaaatgaa	gcttcaaaac	ttgtttcat	ttcagttgt	cacagact	12000
ctcccagg	gactcacacg	tgtaaatatt	cctgactat	tctgcact	ttttatct	12060
tgcttc	ccaaaatgc	aagtgtacaa	ggtgagg	tcaccctt	attcagag	12120
cagggtcg	ctccctaacc	tggacttgc	tttctccggc	agcctctgac	accctcccc	12180
ccat	tatcagaagg	tctgagc	gttggggc	gctcatgt	tgatacact	12240
ttgtcttcc	tgaagatcta	acttctgacc	cagaaagat	gctaaagg	tgaagtgtt	12300
gacatgaaga	cttggctta	agaactgg	cagggaaaaa	aagtccggat	ttgcagc	12360
tacccgaaat	cccagaactg	gggaggt	gacggat	tgccggggc	tagctgg	12420
ctcagccagc	ctagctgaat	tgccaaattc	caactcctat	tgaaaaac	ttaccaaaca	12480
aacaaacaaa	caaataataa	caacaacaac	aacaacaac	tacccat	aagggtggc	12540
gctttggct	cttgaggaat	gactcacca	aacccaa	ttgcac	tgttct	12600
cctaaatgg	gtgggggtgg	ggcagagaca	gagacagaga	gagacat	ttcctgg	12660

gggctgtgtg	ctctaggcca	ccaggaactt	tcctgtcttgc	ctctctgtct	ggcacagcca	12720
gagcaccagc	acccagcagg	tgcacacacc	tccctccgtg	cttcttgagc	aaacacaggt	12780
gccttggtct	gtctattgaa	ccggagtaag	ttcttgcaga	tgtatgcatg	gaaacaacat	12840
tgtcctgggtt	ttatttctac	tgttgcata	aaaaccgggg	aactccagga	agcagctgag	12900
gcagaggcaa	atgcaaggaa	tgcgcctcc	tagcttgcctc	cccatggctt	gccgggcctg	12960
ctttctgcaa	gcccttctct	ccccatttggc	atgcctgaca	tgaacagcgt	ttgaaaatgt	13020
ctcaaatgtc	actttcaaaag	aaggcttctc	tgatcttgct	aactaaatca	gaccatgtt	13080
caccgtgcat	tatctttctg	ctgtctgtct	gtctgtctgt	ctgtctatct	gtctatcatc	13140
tatcaatcat	ctatctatct	atcttctatt	tatctaccta	tcattcaatc	atctatcttc	13200
taactagtta	tcatttattt	atttgtttac	ttactttttt	tatttgagac	agtatttctc	13260
tgagtgacag	ccttggctgt	cctggAACCC	attctgtAAC	caggctgtcc	tcaaactcAC	13320
agagatccaa	ctgcctctgc	ctctctggtg	ctgggggttaa	agacgtgcac	caccaacgccc	13380
cgcctctatc	atctattttat	gtacttattt	ttcagtcatt	atctatcttc	taactatccaa	13440
tcatctgtct	atccatcatc	tatctatcta	tctatctatc	tatctatcta	tctatcatcc	13500
atctataatc	aatttg					13515

SEQ ID NO: 2

14931 bases

DNA

Cricetulus griseus

misc_feature (2176)..(2239)

n is a, c, g, t or nucleotide is missing

catgtacact	tatgcaagta	tgatatggcc	caacacagta	ttttacacca	atttttatct	60
ataaaaatata	catgtacatc	aaaatataatt	attaataata	acatcattat	tctttctttc	120
caagtaataa	acacatacac	tgaaattttg	gttcttgcgg	ataattttaa	tgaaacaggg	180
aatgcaaaatt	tatcttagca	tgttacttc	actttctttg	catagataac	cagtaatcac	240
attgatggat	catgtatgt	aatgtatttt	taggtatcta	aggaattttg	gcttcgtttt	300
gtgcttgcgt	acactgaatt	ctattcccaa	caacagtgtg	taaggattct	gtctgatttc	360
tttttaccagt	atttgtccat	ttgcattttc	tttatttattc	atggctgctg	ttctagaaaag	420
tggaaaggtag	tgtgtcaagt	ctgtttaaca	tgttccctg	atgatcagtg	tcttaacacc	480
tctctgagta	catgttggcc	aatgtcggtt	ctagaccat	ctattcttgc	ttgacttata	540
ctgggtacatg	cctgccaaga	aatttctctt	catcctttct	gtctcttcac	tgattttact	600
gatgtgtgga	tttcacattt	atcatatgga	aatagaagat	acaattttct	ttatttcacag	660
tttggaaagac	tttcaatctc	atagatcattc	attatttttt	gctactgttc	cctatgtat	720
ggtgaaattt	ccatttgaat	aattgtctaa	acaattaaca	agaaagaatc	tatttttact	780
tgcaataact	tccatttcag	aacattttact	acactgttac	tatatccaaa	aactagttt	840
atataatcatg	tgagaaatga	ctaatttata	atttggccat	gacattttt	tcagaaacag	900
aaaaaagtgc	caatacatac	acaatgtat	aaatattaag	acttcagcaa	attaaatatt	960
tatttcatgt	atcacataaa	atttcattat	tatgttttat	ttaaatgtgt	ttttaaaaca	1020
gtgggtatcac	taaatattaa	gttagatgt	tttatgtgt	taatgaattt	atatttttaga	1080
atgttataag	ttgttatata	tcaaataatgt	aataaattttt	attttttagg	tctttctcat	1140
taagggtattt	taattttggg	tcccttttcc	agagtgactc	tagctcatga	tgagttgaca	1200
taaaaaactaa	acgtacaaa	atgtacattt	cattcagtat	tgcacttgc	cttgcactg	1260
aagtttgaat	cagttcatac	atttatgtact	tgggaagttac	attaagctaa	ctttcattgc	1320
tctggcaaaa	tgctcgat	gataagagtc	tattgtggaa	agccatggca	gcaggaaagt	1380
aagactgctg	atgatgtt	atccatagtc	aagacgcaga	aggagatgaa	tgctggat	1440
caacatttt	tgctgttcat	tttctctaga	accctagtc	ataaagatgt	atgacttgc	1500
ttcaaaaatgc	gtcccccctca	gttggcaac	ttttctgtaa	atatccccc	aggcatgtct	1560
agaagattgt	ttcgcaaaata	tttctcaatc	cattcaagtt	gatagtgcag	attaatcact	1620
gcagaataaa	agcctgtac	ttggctcacg	tgccaaaggaa	tatgcacact	cctgacacat	1680
caataaagtta	atcaaagtgt	agcttttgc	tttaacattt	ccagacttat	gtaatgttct	1740
gcacgttctt	cctccatcac	tttttattct	aatgggtttt	ccttgacatt	gaatcacgt	1800

gtggaagctg	cttagaatta	acattgaaaat	ctactgatat	atttatgatg	cagcaattta	1860
gatttactat	tttacttaga	atttttata	attgagagaa	tataatattt	tcacagttat	1920
ctatctgtg	taaatagagg	atttaaaaaa	aaatctctat	aactttttt	tacaacacac	1980
agtaaaaattt	agttaaaattt	taataaagtc	actatgtga	tttcaaagtg	tgctacgccc	2040
acggtgtgtca	cgcaggtgt	gcagaagatg	ccactaagg	gggctaaggc	cgatgggtt	2100
gggtctgcgc	tccctggaga	tgagccccag	gcccgtccct	ggcaatcago	tgcgatcatg	2160
atgcccgtg	agccannnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2220
nnnnnnnnnn	nnnnnnnnnc	tgggtgactt	tatggaaaga	atttgataga	tttcatgatg	2280
tagaagaattt	ttatttaggt	tattttacag	gagactaaga	ccctgggacc	taaagatatc	2340
tgggtcctga	gaatcaggaa	atgggttagag	acgtgggtga	tggtatgaga	cagattttag	2400
agaactctta	gatcatgggc	aatgaccgca	atctgatgct	tagaatagat	catctataaa	2460
caattatgt	gttcttttc	tttctgtgt	atgatctgt	gatgtagccc	ccttgccaaq	2520
ttccctgtatc	ccccttgcca	agttccctga	ttgttaacagt	atataaagcat	tgcttgagag	2580
catattcaac	tacattgagt	gtgtctgtct	gtcatttcct	cgccgattcc	tgatttctcc	2640
ttgagcctt	tcccttgttc	tccctcggtc	ggtggtctcc	acgagaggcg	gtccgtggca	2700
aaagtgtata	aatgttctaa	aacatttgaa	ctctaaaaca	tgcaaaaatga	aaaattaaaa	2760
taaataaaaca	tgaaaattaa	aatatattag	ctgctaaaag	ttaaacaata	ctatataata	2820
ttttgttatt	agaattcaaa	atcacatag	ttggatttaa	tttgaacatt	gcattctttc	2880
aataataattt	tcaataaaaaa	aagtttcccc	atgatagtag	aaaataataa	catatgtatc	2940
tatctatttt	tttaactaca	catatatagc	atttgttca	actaaaataa	atgaatgagc	3000
aaagcaccta	agtaattgg	gtcttattata	tttatgaagc	caatagttt	aaataaaattt	3060
tcatgcataa	ggaggtattt	caaagttaa	acottttttt	aaacagatat	tcccagttac	3120
agaaaattata	atttctaattc	tttcctataa	gtagaatgat	gataattaat	ataggccatt	3180
tgtaaataat	gttcagatta	aaatattctc	tatttcaact	gagaagaatg	atattaaatg	3240
tattatattt	tatttccat	tttgggttgc	ccactattct	atatccctca	gcagtttaaa	3300
tttgggttac	cataatgtgt	tgtgtttgt	tcttaaataat	ggcactaaaa	ttagaataat	3360
ttaatataaa	tctttaggag	aaaagatatt	gaatttatttt	atgttgatag	gaaaatatact	3420
ttaattgtc	caagaataact	ttttcttctt	ttttaggact	gatcagaccc	aggactaata	3480
ttttatatgt	actaattcta	tgtacaaaaa	tatgttatta	tctcatgaat	tctgtctcaa	3540
tattgaggta	ataaaaaatag	tccatcatga	actttaaaat	taaaataatg	attaattaat	3600
ttttatttcat	attttgtttt	tatgaatggt	tatacatcac	atgtgtgcct	ggtgactgt	3660
aatgtcagga	gaaggtatga	aagccactgg	aatttggata	agagataata	tttgagatgt	3720
tatgtgggt	ctgagaattt	gacgcaagcc	atcttcaaga	atagccagca	tactatacc	3780
ctgagtaatc	cattcatccc	tcaataattt	tcttgtaga	cagtaaatat	atttctaaac	3840
tataaatgac	cagaaaaattt	aatgtattat	taatgaagac	attcatctca	tgtgacacac	3900
ttcacctgtc	taaatcagta	acactctctc	cactaattaa	gattttctaa	gtgcatgaca	3960
tttactattt	ctaaagctgt	ccaatggggg	ccagtccttca	gtcagcaccc	agtgagataa	4020
tccatgaatg	catttataatc	tttagaaaaa	ttcttatactt	tgttagtattt	agaacatttt	4080
catgtgaggg	gataaaacaag	gaagcacaga	tgctttctga	tagaaacttt	ctctttaatt	4140
catctagaaa	aaaaaaacct	ctcaggaaaa	tctcttctgc	tctccttccca	atgctctatt	4200
cagcatctc	tccctactta	attctagatc	tttttctcta	tgccttcctt	ctgctgcctt	4260
gctggctctg	ctctatgcct	ccccatgtca	cttttctttt	ctatctcacc	gttaccttct	4320
ctgcctcaact	ctctgccttc	ttctctgttt	ctcacatggc	caggctctgg	acaattatag	4380
ttatatgtt	catttcata	acacatgata	tgtcacatag	tttctctcag	gctagggata	4440
tcacaatgac	tggccaatga	gcaagtggcc	ttgcatgtag	ctctaaatgg	gtgtatggttc	4500
ccagacagta	agtggccatt	tggttgaat	ttgaggttgg	gtgtacatg	aagactgaat	4560
tttcttcaaa	ctctggcctt	gaaatagtaa	aacaacacct	atgaaaatga	cgacctgtat	4620
ttgtctttag	aggcaaccac	atattgtctg	cagggcctgc	tttgaatttg	ctctgaagtt	4680
agcttggttt	tgtaaaagga	agaatccat	atcagcctga	gaaatgtaaa	atatccctagc	4740
atttcaagtc	atcaaaaattt	tatggagagt	ataaatcatc	cttctgacta	ttcatagtca	4800
tatttgtgtc	caccaagtt	aaaacacact	acccaaaggc	tgtggaaaaaa	atgccataaa	4860
ctgttcttat	tagggaggca	tagcagtgg	acctgaggaa	gttacagcaa	caaccagtca	4920
tccagtcaat	aaccccatgg	ctttgccact	tggaggtacc	caataatgtt	tggctttgcc	4980
gagtaggact	ccaacaaaatt	cagagggtca	atttttaaaat	gctgggtgtc	actgctgaac	5040
agtcccattt	ccctctgtatc	aattccacaa	tggaaagctt	tttacactga	ttgccaatca	5100

ttaaacagcc tactcagcat	aaacaggat	gatattattc	tgcatttigt	tacattacta	5160	
gatgaattcc tatttcttcc	tacaata	tagt	gaactgaaaa	aagatacaca	5220	
ccctctacta atcttatgac	ttatatcatt	tcaat	ttca	gaccataatg	5280	
accaaaaacat gtgaagatga	aaaataga	aaa	tgtagaataa	tattacatat	5340	
ggcggactta ttttgttta	tttcttagca	tgcata	gca	tacatgattt	5400	
taataaaggg acaataa	ttcaagaa	ttacccctac	tgaatt	aaaaagaaaa	5460	
ggtcacacat ttactca	aat	at	actggc	aaat	tagacatgaa	5520
aatattgagg taggccttct	gtgaaat	gtc	taaggaa	att	atgtttcata	5580
ccaagtggga atcatatcag	aaagc	agtca	aaag	cattt	ttacaagtaa	5640
gttatatgac ctcccagagc	ttgact	gtct	atacaca	aaa	agtgggttta	5700
aatttggct atgtttttt	aaatgg	cttc	acca	acatg	aaggaaggga	5760
catggatgct tagagattat	gcttcc	cagca	agaaga	attt	agctttggct	5820
aaacatgaca aggtgtgagt	tttatttatt	agaaaattata	taatatttta	agctggggac	5880	
taaaaaattt attgaaacaa	acaggc	aaagg	gata	aggcatg	tactaga	5940
tgtcaatgct gtaatgttat	tttttggacc	aaaat	agtagt	ttcctataga	aatgacaatg	6000
atcttaggtt attattctt	ataa	agatg	caagtt	caca	agatatccta	6060
atcgttttag tcattaata	gagt	gctgt	atagatt	aca	aaaggaa	6120
gagaaataat gata	ttt	tttct	taatt	cttag	aaacattct	6180
tcaatctcag aagcactt	ttg	cttatt	attg	aaacat	atgaaattgt	6240
ttgtctatgg tgacatttca	aaga	acatgt	gac	gtac	agtg	6300
aactgcagct gaatc	ctaa	acttac	atac	at	tctg	6360
gtgtgcacta ccaaaggat	tacta	atgt	cac	gacact	ccctatgt	6420
atcattacat cata	ttt	gttt	tc	tttgc	tttgc	6480
tttccaagca cagagct	taattt	ac	cat	actt	tctgg	6540
caattgaata gaaata	ctacta	attt	tgt	gagac	cc	6600
gataattttt acataaa	tg	cata	caag	taa	gact	6660
gagaaatcag gaaggag	atgg	gag	tt	ggat	gag	6720
ctataatctc aatac	tct	acta	aa	aaa	acat	6780
aacatcagtc ttcaaa	ttt	tttgc	aa	agat	tgt	6840
tatctagaaa acacaaa	aa	aaa	tttgc	tttgc	tttgc	6900
gactacatga aaagt	ttt	tttgc	acta	at	tttgc	6960
ttaatataa taaccat	ttt	tttgc	tttgc	tttgc	tttgc	7020
taataactaa tacc	ttt	tttgc	tttgc	tttgc	tttgc	7080
atttgggaga gagaat	ttt	tttgc	tttgc	tttgc	tttgc	7140
atctctgtat aggt	ttt	tttgc	tttgc	tttgc	tttgc	7200
ttttcaagca aatacc	ttt	tttgc	tttgc	tttgc	tttgc	7260
tattgtt	ttt	tttgc	tttgc	tttgc	tttgc	7320
ttaatagatg ataaa	ttt	tttgc	tttgc	tttgc	tttgc	7380
tgaaaattatg taataa	ttt	tttgc	tttgc	tttgc	tttgc	7440
tacctaaaca ttgaa	ttt	tttgc	tttgc	tttgc	tttgc	7500
ctatatgtat tttat	ttt	tttgc	tttgc	tttgc	tttgc	7560
gagagcatgg gatcatt	ttt	tttgc	tttgc	tttgc	tttgc	7620
gagcaataga aaggat	ttt	tttgc	tttgc	tttgc	tttgc	7680
gaggaggat aactag	ttt	tttgc	tttgc	tttgc	tttgc	7740
gaaacttc	ttt	tttgc	tttgc	tttgc	tttgc	7800
ataatggaaa tatta	ttt	tttgc	tttgc	tttgc	tttgc	7860
agattggaa gcattt	ttt	tttgc	tttgc	tttgc	tttgc	7920
aatgaatatg tcag	ttt	tttgc	tttgc	tttgc	tttgc	7980
ttgaagtgt	ttt	tttgc	tttgc	tttgc	tttgc	8040
tcagagccca tacc	ttt	tttgc	tttgc	tttgc	tttgc	8100
aacagagccca taact	ttt	tttgc	tttgc	tttgc	tttgc	8160
cactacaatg ggg	ttt	tttgc	tttgc	tttgc	tttgc	8220
actctagtaa ctggaa	ttt	tttgc	tttgc	tttgc	tttgc	8280
cacaattatt gtatt	ttt	tttgc	tttgc	tttgc	tttgc	8340
ttgctaaaga aacc	ttt	tttgc	tttgc	tttgc	tttgc	8400

tggaaaacttc atgctttctg	cccagcattc atgttgctgg	aaagagcaat gtacactacc	8460
agtgttagaaa ttaaatcatc	aatcttatca agatgtggat	cctataagtt acaataaaaa	8520
ttagcctgat aagatatccc	caccagaaga atattcacat	aatatgtatg ggagcaacaa	8580
gtattttct aaatttagctt	taatcctatt ctacaagaga	gaatccatat ctagaatagt	8640
tataggatc aagaacccat	ggcttgattt gtcataggcc	caatgggaga tcctaatatt	8700
attgttctac aaaatgaaaa	taactcttaa tgacttgg	ctgcagtaat aagttgtat	8760
gttgctcaac tctcacaaga	gaagtttgtt cttacaataa	atggcaatta aagcagcccc	8820
acaagattta ttcataccg	atctcctcat ggcctatgca	tctagaagct agggaaacaaa	8880
gaggacccta agagagacat	acatggtccc cctggagaag	gggaaggggg caagacctcc	8940
aaagctaatt gggagcatgg	gggaggggag agggagttag	aagaaagaga agggataaa	9000
aggagggaga ggaggacaag	agagagaagg aagatctagt	caagagaaga tagaggagag	9060
caagaaaaga gataccatag	tagagggagc ctgttatgtt	taaatagaaa actggcacta	9120
gggaattgtc caaagatcca	caaggtccaa ctaataatct	aagcaatagt cgagaggcta	9180
ccttaaaagc ctttctctga	taatgagatt gatgactacc	ttatatacca tccttagagcc	9240
ttcatccagt agctgatgga	agcagaagca gacatctaca	gctaaacact gagctagtt	9300
cagacaggga ggagtgtga	gcaaagtcaa gaccaggctg	gagaaacaca cagaaacagc	9360
agacctgaaa aaaatgtgc	acatggaccc cagactgata	gctgggagtc cagcatagga	9420
cttttctaga aaccctgaat	gaggatatac gtttgaggt	ctggtaatc tatgggaca	9480
ctggtagtgg atcaatattt	atccctagtt catgactgga	atttgggtac ccattccaca	9540
tggaggaatt ctctgtcagc	ctagacacat gggggaggtt	ctagttcctg ctccaaataaa	9600
tgtgttagac tttgaagaac	tcccttgaga agactcaccc	tccctgggga gcagaaagg	9660
gatggatgaa ggggttgtga	gggacaggag aggagggag	ggtgagggaa ctgggattga	9720
caagtaatg atgcttgttt	ctaatttaaa tgaataaaagg	aaaagtaaaa gaagaaaaga	9780
aaacaggcga aaagattata	aaagacagag gtggtggtg	actataaaga aacactatta	9840
tctaaataaa aataatgtcag	aagcacacat qaacttataq	tgtttatgaa agtatgtata	9900
ataactacat aatctcaagc	caagaaaaaaa atatcatctt	tcagtgtatga aggtgatttt	9960
atttctccca gaattaaagc	caaagaccta atgaaagtaa	ttatcttcaa aaggtgaaa	10020
atacatactt tgcaatacac	agatctgcct agaaaatctca	tgttcacaat acacatgat	10080
ctcaattgaa ttccattcaa	tgttacagtt tagataaaca	gttttagat aaactcacaa	10140
tgtatcattt ctttttat	tttgacaaa cagttctca	tctgttattc agaataattc	10200
ctcgatggca ggatatccat	cccaatttgg ggaagggag	aatttgaaga aaacctagac	10260
cacatacata tttgccattt	ggaaacaaaag tctaaaatga	tgttggtcac atttctcta	10320
ctagtcctc ccccgccca	aagaacctt gatatgtgc	ctcattttac agagagagga	10380
aagcaggaac tgagcatccc	ttacttgcca tcctcaaccc	aaaatttgc tcattgctca	10440
gtctgcct tctcatatga	cagttacaag tcaaggcttc	caaagtccct ctgtcatgtt	10500
tgggtgtcaat agtttataca	gatgacttca tgcattcata	tctaattgtct tatatagatt	10560
aatattaaaac aatgttattt	ctcttaaccac attttaaatt	aatttaaaaa tccattaatt	10620
gtgtctataa aatgcagaca	gagtgtcgag acacaatata	agcctgatga tctgaatttt	10680
aaactcacac ccaccacatg	gagaatcaac ttccaaaaat	tttccttatta ctccacact	10740
tacaccatg tacaaacaca	ataataatga aaaaaatgaa	atgaaataaa aaattaagtc	10800
tctgttaggtt atgctactgt	gcagcaaaag taaaaatggc	agcttaagct tgctttatgg	10860
ttacacttta ccatcttcca	ttaattataa ggacttcaat	catggcagaa ctatgctgtt	10920
attgtctcag tgcattacaa	ccaggtgttc cagatgttct	taatgtggac acctaaacta	10980
tttgcattttt gggtaagat	ctttccctt ttcagaagaa	acctcaggac agagggatc	11040
ttgtctttt attttgagtc	tgcatttttca aatatacatg	aaacaagtga ttaagcttca	11100
tgaagaaaat taatcaaaaag	gtgggatttgaat	caatgtatatt aggttcaata	11160
atattatcat ggaatcgcc	gttatacact gatgtttgg	caataaggaa tttttagaaag	11220
aaggagtttt tatttctcaac	agttcccttta agtttagctc	aaataaaatct aagcaatcca	11280
ctctagaatt aaatagttc	ctaaggccac agctatgaat	agagctcaat ttacatataa	11340
aattttgttc accattatg	tcattccagt tttcatttagt	acaaggaaaa tacaaaatat	11400
ttagatgtca atatcaagtg	aatagttcat ctccctttt	aatataatct acctaaatca	11460
ccatccatc agaaaaatct	ggcctgaagt tctgtctgg	acttcaacat gaaaaatatg	11520
cacagcttgc tattataat	cctagttgtat tttttagatt	catgtctggt gtctgactca	11580
gagggccag aggctagaca	aatattttt gaatcttcat	tgtgaagatt tttatgatt	11640
attttaat aataacaacaa	gatgtggat aatgtactt	tgtacagtt atagacgctg	11700

aactactttg	tgcttaaaat	gttagttccc	tatcataaaat	gatagggtat	aagtgtatgt	11760
ttaataacttt	ccctctgagc	tatattcatg	tactagagaa	ttattttaaa	catgaaaaga	11820
ctgtgtttat	agtctcagct	cctgagaact	ggtccaaccc	taggcagggt	aatgccagg	11880
gcaacgtttt	tcttctacag	aggatgcctt	gctgccaagc	aacctgggt	tgtggaaatg	11940
ttcccttttt	aatcaagttt	aaagggtctt	catcatgctg	ttgctccaca	tatttcagg	12000
ttagagcttg	gtccttggag	tattatctt	taccagaaaa	ttcatagtat	tctttcaata	12060
actaaacaact	aaacttttcg	ataaaaaaaga	atttggaaattt	caattttaaa	gcctgagtaa	12120
aattcttgcg	aatcaggata	ttttattttt	agtcttatct	tttaaaaaagt	tattttattt	12180
tttaaaaaat	tataatatac	tttcataatt	tcocctccttc	acttttctt	acaaacactt	12240
ctataagatca	ccatgtgttt	ttttttttac	atttatggcc	tctttctgtt	cattgttatt	12300
acatacaaaat	agtcttgcct	atagaagaac	accacaattt	gttacctgat	aacaaattat	12360
caacccttaa	aacctacaaa	ctattgat	tactgaaaag	actataactt	tagatgtaaa	12420
gatatatgtg	tgtgcacata	tatagataca	catatatgt	ggatttttaa	tttttagattt	12480
tagacatcaa	aatttatttt	atgactgaga	aactagacac	tataaatgag	cattcagtat	12540
tcaacaccgt	gatttttagat	attgtcacaa	tgacagaaaa	ttttcttata	gaaaatttta	12600
agttttgtga	ttgcctgtg	cacttagtga	agtctcacag	aaaaagaatc	atagtatttt	12660
tagtttataa	taaaaaagtac	atataattaa	aatgggtggc	acaaaacaaac	atttgagcat	12720
ttttccattt	tactatcaag	tagtattcatt	ttgaaataat	aatttgacta	gtttcaaaaa	12780
tgaaaaacaaa	atttaaacta	aatgcctaat	ctagcctgat	aacatttttta	tgaatgaaat	12840
tattcaatag	tgtttatcaat	tagggccca	aaacctttcc	taaaataaaaa	cttttaattt	12900
ttttccattt	ttattttaaat	tagaaacaaa	attgttttac	atgtaaatca	gagtttcctc	12960
acccctccct	tctccctgtc	cctcacta	accctactt	tcccatacc	tttctgctcc	13020
ccagggaggg	tgaggccttc	catggggaaa	cttcagagtc	tgtctatcct	ttcggatagg	13080
gcctaggccc	tcacccattt	gtctaggcta	aggctcacaa	agtttactcc	tatgcttagt	13140
ataagtactg	atctactaca	agagacacca	tagatttcc	aggcttcctc	actqacaccc	13200
atgttcatgg	ggtctggaac	aatcatatgc	tagtttcc	ggtatcagtc	tgcccaccat	13260
gagctccccc	ttgttcaggt	caactgtt	tggtgggttc	accaccctgg	tcttgcactgc	13320
tttgctcatc	actccctccct	ttctgttaact	gggttccagt	acaattccgt	gtttagctgt	13380
gggtgtctac	ttctacttcc	atcagctt	gggatggagc	ctctaggata	gcatacaatt	13440
agtcatcatc	tcattatcag	ggaaggccat	ttaaagttagc	cttcatttgc	ttgcttggat	13500
tgtagttgg	tgtcatctt	gtagatct	ggacatttcc	ctagtgccag	atatcttctt	13560
aaacacctaa	gactacctct	attatgtat	ctcttttctt	gctctcgct	attcttccag	13620
acaaaatctt	cctgctccct	tatatttcc	tctccctcc	tcttctcccc	ttctcattt	13680
cctagatcca	tcttcccttc	ccccatgctc	ccaagagaga	tggtgcctag	gagatcttgc	13740
tccttaaccc	ttttcttggg	gatctgtctc	tcttagggtt	gtccttgc	cctagcttct	13800
ctggaagtgt	ggattgtaa	ctggtaatca	tttgcctccat	gtctaaaatc	catastatgag	13860
tgatgtttgt	cttttgcgt	ctgggttacc	tcactcaaaa	tggtttctt	cataatgtct	13920
tggatttcaa	tagcacaaac	aacatacagt	atcttggggc	aacactaacc	aaacaagtga	13980
aagaccagta	tagcaagaac	tttgagttt	aagaaagaaaa	ttaaagaaga	taccagaaaa	14040
tggaaagatc	tcccatgctc	tttgataggc	agaatcaaca	tagaaaaat	ggcaatcttgc	14100
ccaaaatcca	tctacagact	caatgcaatc	cccattaaat	accagcacac	ttcttcacag	14160
acctgaaaga	ataataactt	actttatatg	gagaaacaaa	agacccagg	tagccaaac	14220
aaccctgtac	aatgaaggca	cttccagagg	cattccatc	cctgacttca	agctcttatt	14280
tagagtaata	atccgtaaaa	cagcttgcgt	atggcacaaa	aatagacagg	tagaccaatg	14340
gaattgagtt	gaaaaccctg	atattaaccc	acatatctat	gaacacccct	ctttgacaaa	14400
gaagctaa	ttatataatgt	taagaaagaa	agcatcttca	acaaatcg	ctggcataac	14460
tggatgtcg	catgtagaag	actgcagata	gatccatgtc	taatgccc	cacaaaactt	14520
aagtccaaat	ggatcaaaaa	cctcaacata	aatccagcc	cactgaaccc	catagaagag	14580
aaagtgggaa	gtatccctga	ataaaattgt	acaggagacc	acatcttgc	cttaacacca	14640
gtacgcacaga	caatcagatc	aataatcaat	aatggggacc	tcctgaaact	gagaagctt	14700
tgtaggccaa	tggataagtc	aacaggacaa	aatggcagcc	cacggatgg	gaaaagat	14760
tcaccaatcc	tatatactgac	agagggctgc	tcttatttgc	caaagaacac	aataagctag	14820
tttttaaaac	accaattaat	ccgattataa	agtgggtag	agaactaaat	aaagaattgt	14880
taacagagca	atctaacttgc	gcagaaagac	acataagaaa	gtgctcacca	t	14931

SEQ ID NO: 3

4001 bases

DNA

Cricetulus griseus

ccaagatgcc	catcaactga	ttaatagatg	ataaaattat	tgtacatttc	agtgtaatat	60
tattcagtt	ttaagaaaaaa	tgaaattatg	taataagcat	gtaaatggat	atatcttcaa	120
acaaccatc	cccattatat	tacctaaaca	ttgaaagtcc	aaaatcatat	gatctttta	180
gtggatctac	taatctttg	ctatatgtat	tttattgaac	tacccatgga	tgtgagataa	240
ttggtaacaa	cagcacatgg	gagagcatgg	gatcattcaa	ggaagattag	agagaatgca	300
tttttagga	gataatggag	gagcaataga	aaggattaaa	tgaggttact	gatgaaagtg	360
atggtagag	aaggcaatat	gaggaggat	aactagcact	tagggcctt	tgaaaaagac	420
atagagaaaa	tactattgt	gaaacttcct	ataatttggtg	tatagttata	tacaccaaag	480
agctcagatg	gagttaccct	ataatggaaa	tattaactac	tttttatcac	tgtgataaaa	540
catcctgaac	agagcaacat	agattggaa	gcatttactt	tggcttacag	ttctaaccggg	600
ataaaaatc	atgatgaaag	aatgaatatg	tcagcaaaca	gcagtagcaa	tggcttgaga	660
agcaggtgag	agctcacatc	ttgaagtgt	agaatgtac	agagagaaca	aactgcaaat	720
gaccagaaaa	tgcttttgg	ttagagccca	tacccctctg	actgacttct	ccagaaaattc	780
tgaacaaata	aaactcccc	aacagagcca	taactgaagg	tccagttct	gagactacta	840
ggggtatttc	ttattcaaac	cactacaatg	gggtgggggg	agcaatcctc	caagtaggca	900
ctacacacag	acaataaaaa	actctagtaa	ctggaatgga	ttgacttatt	tgaattactt	960
gccagtggag	ctacatagag	cacaattatt	gtatttaat	taccctttat	gatcttacaa	1020
aacttgacag	taagatcata	ttgctaaaga	aaccacatat	ttgaatcagg	gaacatggtg	1080
atatctagtt	gttcttcaac	tggaaacttc	atgctttctg	cccagcattc	atgttgcgg	1140
aaagagcaat	gtacactacc	agtgtagaaa	ttaaatcatc	aatcttatca	agatgtggat	1200
cctataagtt	acaataaaaa	ttagcctgat	aagatatccc	caccagaaga	atattcacat	1260
aaatgctatg	ggagcaacaa	gctattttct	aaatttagctt	taatcctatt	ctacaagaga	1320
gaatccatat	ctagaatagt	tatagggtac	aagaacccat	ggcttgattt	gtcataggcc	1380
caatgggaga	tcctaatatt	attgttctac	aaaatgaaaa	taactcctaa	tgacttgcgg	1440
ctgcagtaat	aagttagtat	tttgctcaac	tctcacaaga	gaagttttgt	cttacaataaa	1500
atggcaat	aagcagcccc	acaagattt	tatcataccg	atctcctcat	ggcctatgca	1560
tctagaagct	aggaaacaaa	gaggacccta	agagagacat	acatggtccc	cctggagaag	1620
gggaaggggg	caagacctcc	aaagctaatt	gggagcatgg	gggagggggag	agggagtttag	1680
aagaaagaga	aggggataaa	aggagggaga	ggagggacaag	agagagaagg	aatgatctagt	1740
caagagaaga	tagaggagag	caagaaaaga	gataccatag	tagagggagc	tttgcgtatgtt	1800
taaatagaaa	actggcacta	ggaaatttgc	caaagatcca	caaggtccaa	ctaataatct	1860
aagcaatagt	cgagaggcta	ccttaaaagc	ctttctctga	taatgagatt	gatgactacc	1920
ttatatacc	tccttagagcc	ttcatccagt	agctgtatgg	agcagaagca	gacatctaca	1980
gctaaacact	gagctagttg	cagacaggg	ggagtgtatga	gcaaagtcaa	gaccaggctg	2040
gagaaacaca	cagaaacacgc	agacctgaaa	aaaatgttgc	acatggaccc	cagactgata	2100
gctgggagtc	cagcatagga	cttttctaga	aaccctgaat	gaggatatac	gtttggaggt	2160
ctggtaat	tatggggaca	ctggtagtgg	atcaatattt	atcccttagtt	catgactgga	2220
atttgggtac	ccattccaca	tggaggaatt	ctctgtcagc	ctagacacat	gggggaggtt	2280
ctaggtcc	ctccaaataa	tgtgttagac	tttgaagaac	tcccttgaga	agactcaccc	2340
tccctggg	gcagaaagggg	gatgggatga	gggttggtga	gggacaggag	aggagggggag	2400
ggtgagggaa	ctgggattga	caagtaatg	atgcttgcgtt	ctaatttaaa	tgaataaagg	2460
aaaagtaaaa	gaagaaaaga	aaacaggcca	aaagattata	aaagacagag	gtggtggtt	2520
actataaaga	aacactatta	tctaaataaa	aacatgtcag	aagcacacat	gaactttag	2580
tgtttatgaa	agtatgtata	ataactacat	aatctcaagc	caagaaaaaa	atatcatctt	2640
tcagtgtat	aggtagattt	atttctccca	gaattaaagc	caaagaccta	atgaaaagtaa	2700
ttatcttcaa	aagggttggaa	atacataactt	tgcaatacac	agatctgcct	agaaatctca	2760
tgttcacaat	acacatgtat	ctcaatttggaa	ttccattcaa	tgttacagtt	tagataaaaca	2820
gtttgtat	aaactcaca	tgtatcatt	cttttattt	tttgacccaaa	cagttctca	2880
tctgttattc	agaataattc	ctcgatggca	ggatattccat	cccaattggg	ggaagggggag	2940
aatttgaaga	aaacctagac	cacatacata	tttgcatttgc	ggaaacaaaag	tctaaaatga	3000

tgttgttac	atcttctcta	ctagtcctct	ccccgtcccc	aagaaccitg	gtatatgtgc	3060
ctcattttac	agagagagga	aagcaggaac	ttagcatccc	ttacttgcca	tcctcaaccc	3120
aaaatttgc	tcattgtca	gctctgcctt	tctcatatga	cagttacaag	tcaaggcttc	3180
caaagtcct	ctgtcatgtt	tggtgtcaat	agtttataca	gatgacttca	tgtcttcata	3240
tctaattgtct	tatatagatt	aatattaaac	aatgttattt	ctctaaccac	attttaaatt	3300
aatttaaaaa	tccattaatt	gtgtctataa	aatgcagaca	gagtgctgag	acacaatata	3360
agcctgatga	tctgaatttg	aaactcacac	ccaccacatg	gagaatcaac	ttccaaaaat	3420
tttccttata	cttccacact	tacaccattg	tacaaacaca	ataataatga	acaaaatgaa	3480
atgaaataaa	aaattaagtc	tctgttaggt	atgctactgt	gcagcaaaag	taaaaatggc	3540
agcttaagct	tgctttatgg	ttacactta	ccatcttcca	ttaattataa	ggacttcaat	3600
catggcagaa	ctatgtgttt	attgtctcaq	tgttaacctaa	ccagggtttc	cagatgttct	3660
taatgtggac	acctaaacta	tttgatattt	gggttaagat	ctttccctct	ttcagaagaa	3720
acctcaggac	agagggaaatc	ttgtctttta	attttgagtc	tgttagacttt	ttccatttca	3780
aatatacatg	aaacaagtga	tgaagaaaaat	taatcaaaag	gtgggaatttgc	caatgatatt	3840
agggtcaata	ttaagcttca	atattatcat	ggaatcgctt	gttatacact	gagtgtttgg	3900
caataaggga	tttttagaag	aaggagtttt	tattctcaac	aggttcctta	agtttagctc	3960
aaataaatct	aagcaatcca	ctctagaatt	aaatagttc	c		4001

WHAT IS CLAIMED IS:

1. A cell comprising an exogenous nucleic acid sequence integrated at a specific site within an enhanced expression locus, wherein the exogenous nucleic acid sequence encodes a bispecific antigen-binding protein.
2. The cell of claim 1, wherein the exogenous nucleic acid sequence comprises a first exogenous nucleic acid comprising a nucleotide sequence encoding a first LCF, a second exogenous nucleic acid comprising a nucleotide sequence encoding a first HCF, and a third exogenous nucleic acid comprising a nucleotide sequence encoding a second HCF.
3. The cell of claim 2, wherein the nucleotide sequence encoding the first HCF encodes a first CH2 and a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH2 and second CH3 domain.
4. The cell of claim 3, wherein the first and second CH3 domains differ in at least one amino acid position.
5. The cell of claim 3, wherein the first and second CH3 domains are CH3 domains of a human IgG, and one of the two CH3 domains have been modified in at least one amino acid position resulting in a Protein A binding characteristic different from the unmodified CH3 domain.
6. The cell of claim 3, wherein the nucleotide sequences encoding the first and second CH domains differ from each other in that one of the nucleotide sequences has been codon modified.
7. The cell of claim 2, wherein the nucleotide sequence encoding the first HCF encodes a first VH, and the nucleotide sequence encoding the second HCF encodes a second VH.
8. The cell of claim 2, wherein the nucleotide sequence encoding the first LCF encodes a first VL.
9. The cell of claim 1, wherein the exogenous nucleic acid sequence further comprises an additional exogenous nucleic acid comprising a nucleotide sequence encoding a second LCF.

10. The cell of claim 9, wherein the nucleotide sequence encoding the second LCF encodes a second VL.

11. The cell of claim 2, wherein the first exogenous nucleic acid further comprises a first promoter operably linked to the nucleotide sequence encoding a first LCF, the second exogenous nucleic acid further comprises a second promoter operably linked to the nucleotide sequence encoding a first HCF, and a third exogenous nucleic acid comprising a third promoter operably linked to the nucleotide sequence encoding a second HCF, wherein the first, second and third promoters are the same or different.

12. The cell of claim 2, wherein the first exogenous nucleic acid is located upstream relative to the second and third exogenous nucleic acids.

13. The cell of claim 12, further comprising a first recombinase recognition site (RRS) located 5' relative to the first exogenous nucleic acid, and a second recombinase recognition site (RRS) located 3' relative to both the second and third exogenous nucleic acids, wherein the first and second RRSs are different.

14. The cell of claim 12, further comprising a third RRS, located 3' relative to the first exogenous nucleic acid, and 5' relative to one or both the second and third exogenous nucleic acids, wherein the third RRS is different from the first and second RRSs.

15. The cell of claim 13, further comprising a fourth exogenous nucleic acid comprising a selectable marker gene.

16. The cell of claim 15, wherein the fourth exogenous nucleic further comprises a third RRS and is located 3' relative to the first exogenous nucleic acid.

17. The cell of claim 2, wherein the second exogenous nucleic acid is located upstream relative to the first and third exogenous nucleic acids.

18. The cell of claim 16, wherein the order of the exogenous nucleic acids at the locus is: from 5' to 3', the first exogenous nucleic acid, the fourth exogenous nucleic acid, the second exogenous nucleic acid, and the third exogenous nucleic acid, wherein the second exogenous nucleic acid comprises a nucleotide sequence encoding a modified CH3 domain of a human

IgG, and the third exogenous nucleic acid comprises a nucleotide sequence encoding the native CH3 domain of the human IgG.

19. The cell of claim 16, wherein the order of the exogenous nucleic acids at the locus is: from 5' to 3', the first exogenous nucleic acid, the second exogenous nucleic acid, the fourth exogenous nucleic acid, and the third exogenous nucleic acid, wherein the second exogenous nucleic acid comprises a nucleotide sequence encoding the native CH3 domain of a human IgG, and the third exogenous nucleic acid comprises a nucleotide sequence encoding a modified CH3 domain of the human IgG.

20. The cell of claim 18 or 19, wherein the first, second and third promoters are the same and are different from the promoter to which the selectable marker gene is operably linked.

21. The cell of claim 1, wherein the bispecific antigen binding protein specifically binds to a T-cell antigen and to a tumor cell antigen.

22. The cell of claim 1, wherein the enhanced expression locus is selected from the group consisting of a locus comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1, and a locus comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 2.

23. The cell of any of the preceding claims, wherein the cell is a CHO cell.

24. A set of vectors, comprising

a first vector comprising from 5' to 3': a first RRS, a first nucleic acid comprising a nucleotide sequence encoding a first LCF, and a third RRS;

a second vector comprising from 5' to 3', said third RRS, a second nucleic acid comprising a nucleotide sequence encoding a first HCF, a second RRS;

wherein either the first or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF; and

wherein the first and second HCFs, and the first LCF, encode a bispecific antigen-binding protein.

25. The set of vectors of claim 24, wherein the nucleotide sequence encoding the second HCF is included in the first nucleic acid.

26. The set of vectors of claim 25, wherein the nucleotide sequence encoding the second HCF is located upstream or downstream of the nucleotide sequence encoding the first LCF.
27. The set of vectors of claim 24, wherein the nucleotide sequence encoding the second HCF is included in the second nucleic acid.
28. The set of vectors of claim 24, wherein the nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain.
29. The set of vectors of claim 28, wherein the first and second CH3 domains differ in at least one amino acid position.
30. The set of vectors of claim 28, wherein the first and second CH3 domains are CH3 domains of a human IgG, and one of the two CH3 domains have been modified in at least one amino acid position resulting in a Protein A binding characteristic different from the unmodified CH3 domain.
31. The set of vectors of claim 28, wherein the nucleotide sequences encoding the first and second CH domains differ from each other in that one of the nucleotide sequences has been codon modified.
32. The set of vectors of claim 24, wherein the nucleotide sequence encoding the first HCF encodes a first VH, and the nucleotide sequence encoding the second HCF encodes a second heavy chain VH.
33. The set of vectors of claim 24, wherein the nucleotide sequence encoding the first LCF encodes a first VL.
34. The set of vectors of claim 24, wherein the nucleotide sequence encoding a first light chain LCF is operably linked to a first promoter, the nucleotide sequence encoding a first HCF is operably linked to a second promoter, and the nucleotide sequence encoding a second HCF is operably linked to a third promoter.
35. The set of vectors of claim 34, wherein the first, second and third promoters are the same or different.

36. The set of vectors of claim 24, wherein the first nucleic acid further comprises a second promoter located 5' to the third RRS in the first vector; and the second nucleic acid further comprises the selectable marker gene, located 3' to the third RRS in the second vector.

37. The set of vectors of claim 24, wherein the first nucleic acid further comprises a second promoter in the first vector located 5' to a 5' portion of a selectable marker gene; and a 3' portion of the selectable marker in the second vector is present 3' to a homologous arm of the 5' portion of the selectable marker gene.

38. The set of vectors of the claim 24, wherein

the first vector comprising from 5' to 3', the first RRS, the first nucleic acid, and the third RRS; and

the second vector comprising 5' to 3', the third RRS, the second nucleic acid wherein the second nucleic acid comprises the nucleotide sequence encoding a first heavy chain or a fragment thereof and the nucleotide sequence encoding a second heavy chain or a fragment thereof.

39. The set of vectors of the claim 24, wherein

the first vector comprising from 5' to 3', the first RRS, the first nucleic acid wherein the first nucleic acid comprises the nucleotide sequence encoding a first LCF and the nucleotide sequence encoding a second HCF, and the third RRS; and

the second vector comprising 5' to 3', the third RRS, the second nucleic acid wherein the second nucleic acid comprises a nucleotide sequence encoding a first HCF.

40. The set of vectors of claim 38 or 39, wherein the first nucleic acid further comprises a 5' portion of a selectable marker gene, located at 5' to the third RRS in the first vector; and the second nucleic acid further comprises the remaining 3' portion of the selectable marker gene, located 3' to the third RRS in the second vector; and wherein optionally the third RRS in the first vector is present within a 5' portion of an intron of the selectable marker gene, and the third RRS in the second vector is present within a 3' portion of an intron of the selectable marker gene.

41. The set of vectors of claim 24, further comprising a third vector comprising one or more RRSs and a nucleotide sequence encoding a second LCF.
42. The set of vectors of claim 24, further comprising a third vector encoding one or more recombinase that recognize the RRSs.
43. A vector, comprising an exogenous nucleic acid sequence that encodes a bispecific antigen-binding protein, flanked by a 5' homology arm and a 3' homology arm for integration into an expression enhancing locus of a cell.
44. A system comprising a cell and a set of vectors,
 - wherein the cell comprises, integrated within an enhanced expression locus of its genome from 5' to 3': a first RRS, a first exogenous nucleic acid, a third RRS, a second exogenous nucleic acid, and a second RRS, wherein the three RRSs are different from one another;
 - wherein the set of vectors comprises
 - a first vector comprising from 5' to 3', the first RRS, a first nucleic acid comprising a nucleotide sequence encoding a first LCF, and the third RRS;
 - a second vector comprising the third RRS, a second nucleic acid comprising a nucleotide sequence encoding a first HCF, and the second RRS; and
 - wherein either the first nucleic acid or the second nucleic acid further comprises a nucleotide sequence encoding a second HCF; and
 - wherein upon introduction of the vectors into the cell, the first and second nucleic acids in the vectors integrate into the enhanced expression locus through recombination mediated by the first, second and third RRSs.
45. The system of claim 44, wherein the first exogenous nucleic acid comprises a first selectable marker gene, and the second exogenous nucleic acid comprises a second selectable marker gene, wherein the first and second selectable marker genes are different.
46. The system of the claim 44, wherein

the first vector comprising from 5' to 3', the first RRS, the first nucleic acid comprising the first LCF, and the third RRS; and

the second vector comprising 5' to 3', the third RRS, the second nucleic acid, wherein the second nucleic acid comprising both the nucleotide sequence encoding the first HCF and the nucleotide sequence encoding the second HCF, and the second RRS.

47. The system of the claim 44, wherein

the first vector comprising from 5' to 3', the first RRS, the first nucleic acid comprising the nucleotide sequence encoding the first LCF and the nucleotide sequence encoding the second HCF, and the third RRS; and

the second vector comprising 5' to 3', the third RRS, the second nucleic acid comprising the nucleotide sequence encoding the first HCF, and the second RRS.

48. The system of claim 46 or 47, wherein the first nucleic acid in the first vector further comprises a 5' portion of a selectable marker gene, located at 5' to the third RRS; and the second nucleic acid in the second vector further comprises the remaining 3' portion of the selectable marker gene, located 3' to the third RRS.

49. The system of claim 48, wherein the third RRS in the first vector is present within a 5' portion of an intron of the selectable marker gene; and the third RRS in the second vector is present within a 3' portion of an intron of the selectable marker gene.

50. The system of claim 48, wherein the nucleotide sequence encoding the LCF is operably linked to a first promoter, the nucleotide sequence encoding the first HCF is operably linked to a second promoter, and the nucleotide sequence encoding the second HCF is operably linked to a third promoter, wherein the first, second, and third promoters are the same or different, and/or said promoters are the same or different from the promoter to which the selectable marker gene is operably linked.

51. The system of claim 44, wherein the nucleotide sequence encoding the first HCF encodes a first CH3 domain, and the nucleotide sequence encoding the second HCF encodes a second CH3 domain.

52. The system of claim 51, wherein one of the first and second CH3 domains is the CH3 domain of a human IgG, and the other one is a modified CH3 domain of the human IgG comprising a modification of at least one amino acid position.

53. The system of claim 52, wherein the nucleotide sequence encoding the modified CH3 domain is in the first vector.

54. The system of claim 52, wherein the nucleotide sequence encoding the modified CH3 domain is in the second vector and is upstream of the nucleotide sequence encoding the unmodified CH3 domain.

55. A method, comprising:

(i) providing a system according to any one of claims 44-54;

(ii) introducing the vectors simultaneously into the cell by transfection; and

(iii) selecting a transfected cell where the first and second nucleic acids in the vectors have integrated into the enhanced expression locus of the cell through recombination mediated by the first, second and third RSSs.

56. The method of claim 55, further comprising:

(iv) expressing the first LCF, the first HCF, and the second HCF in the selected transfected cell; and

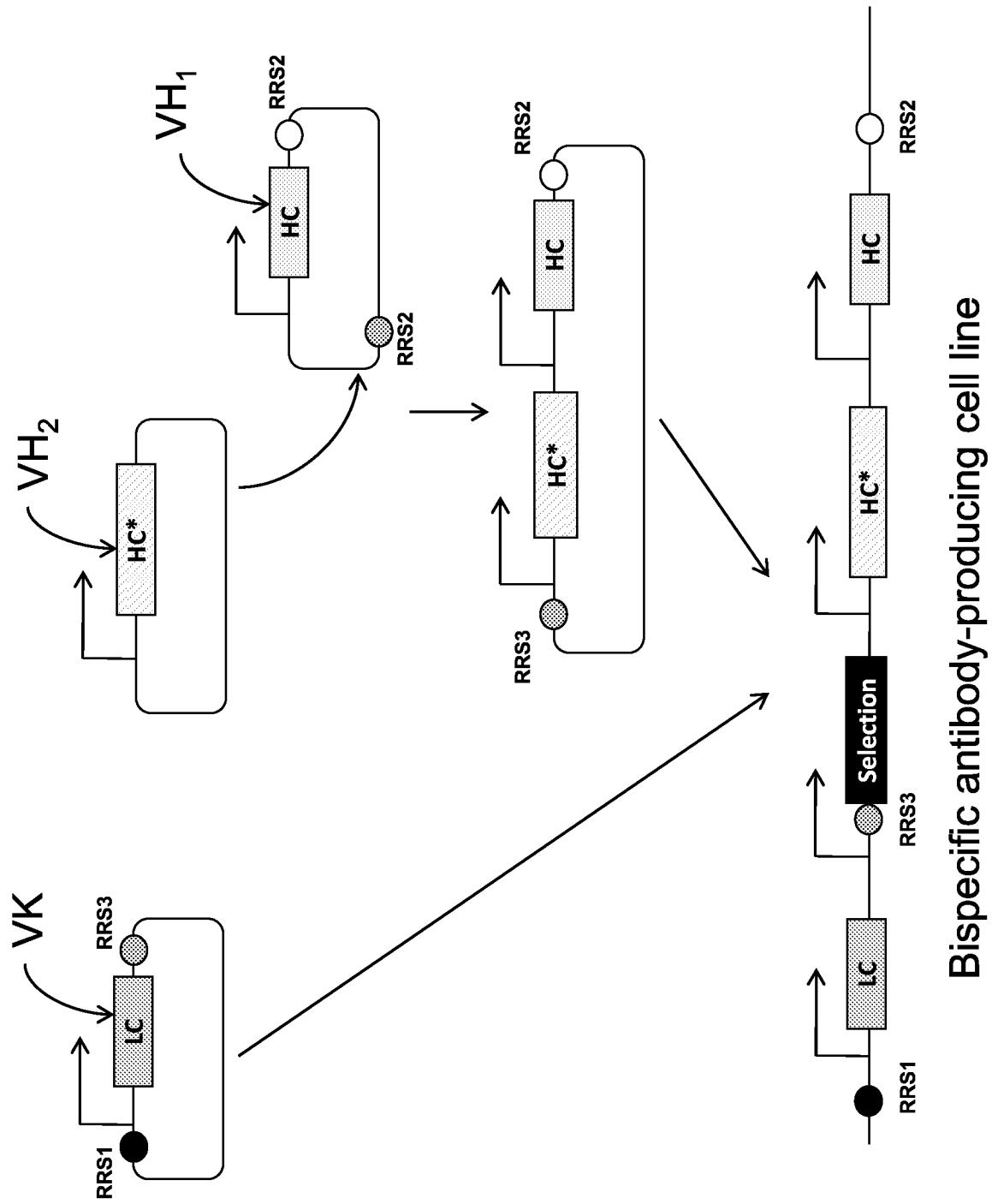
(v) obtaining the bispecific antigen-binding protein comprising the first LCF, the first HCF and the second HCF from the selected transfected cell.

57. A method of making a bispecific antigen-binding protein, comprising:

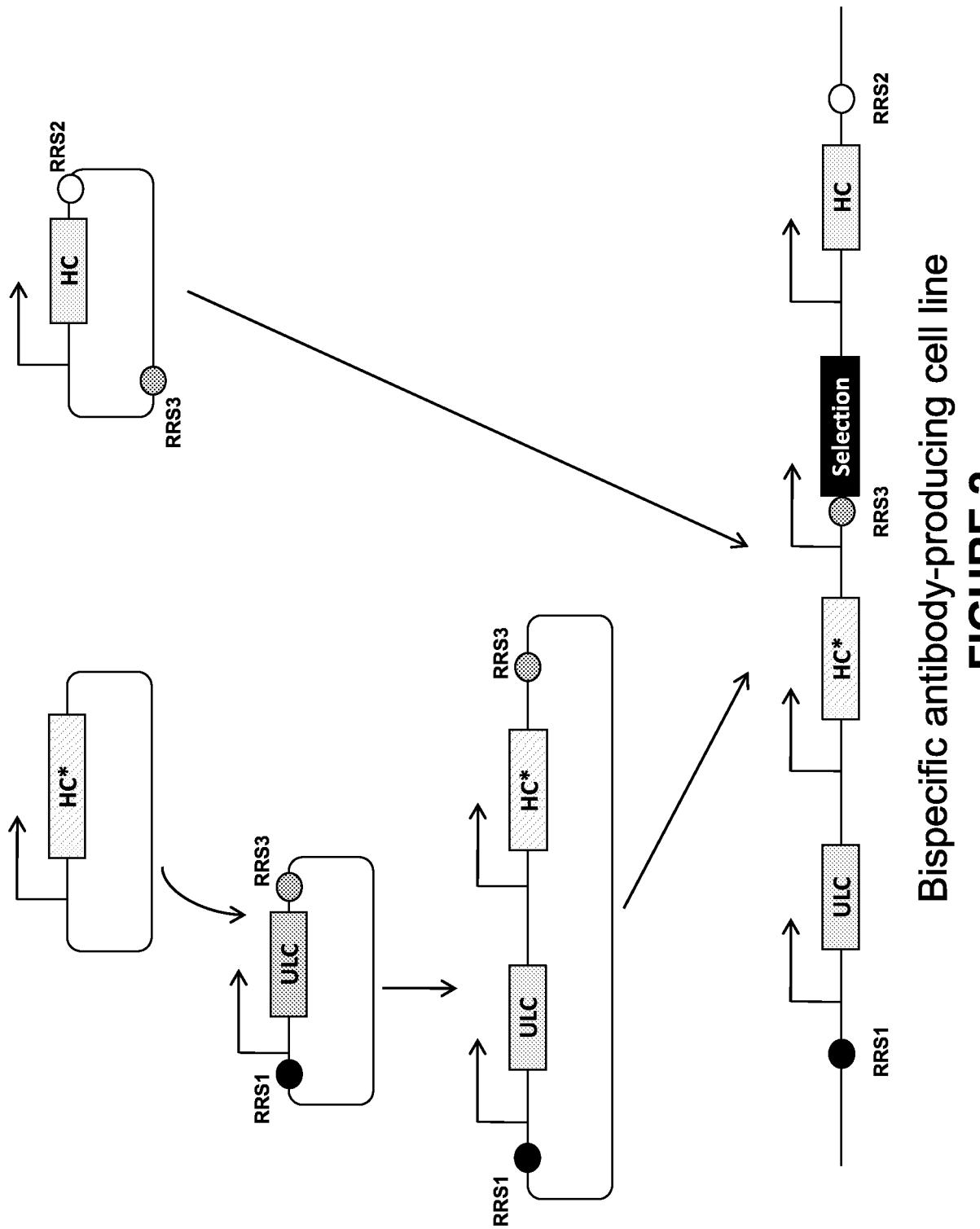
(i) providing the cell according to any one of claims 1-23;

(ii) expressing the bispecific antigen-binding protein from the exogenous nucleic acid sequence; and

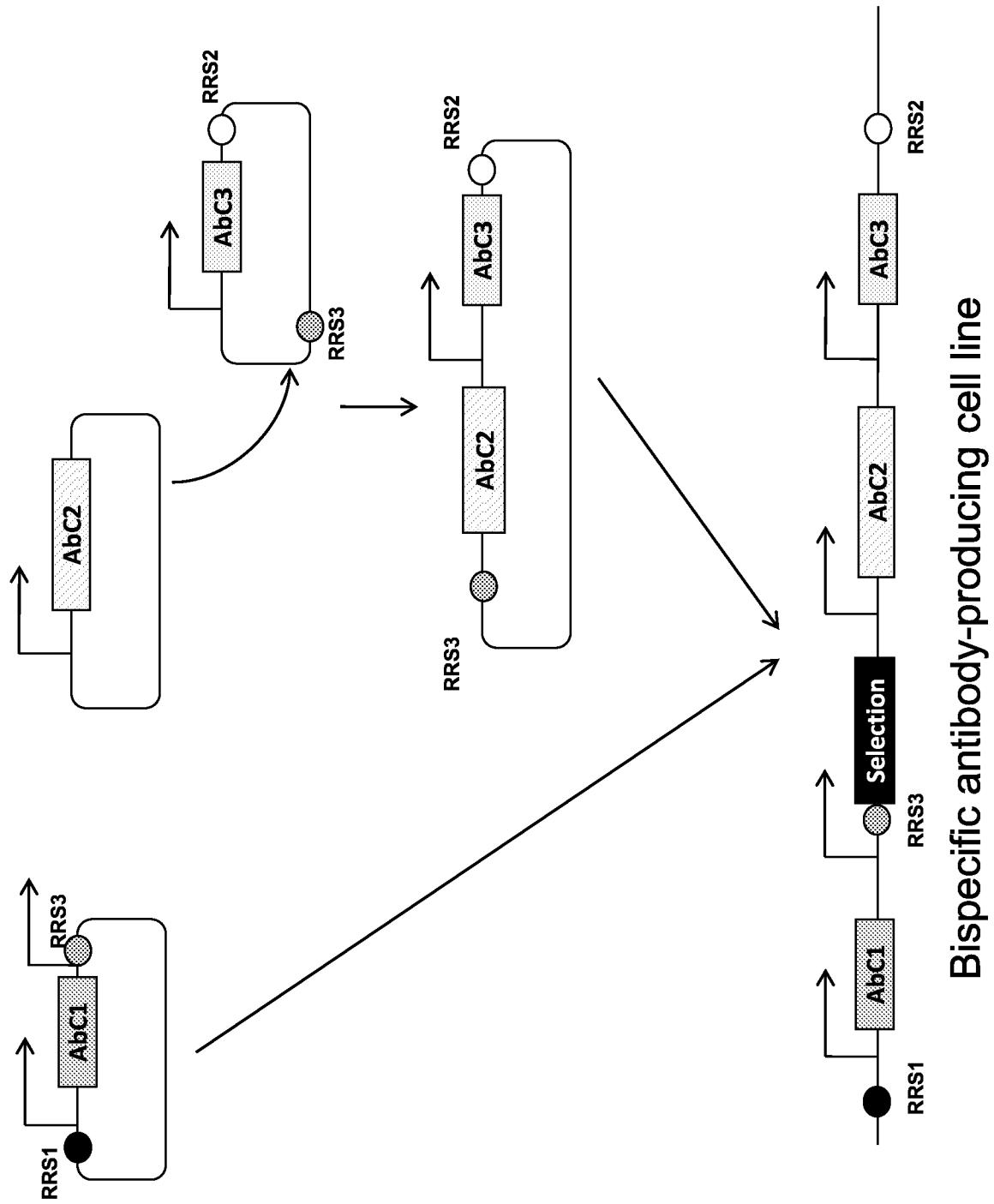
(iii) obtaining the bispecific antigen-binding protein from the cell.

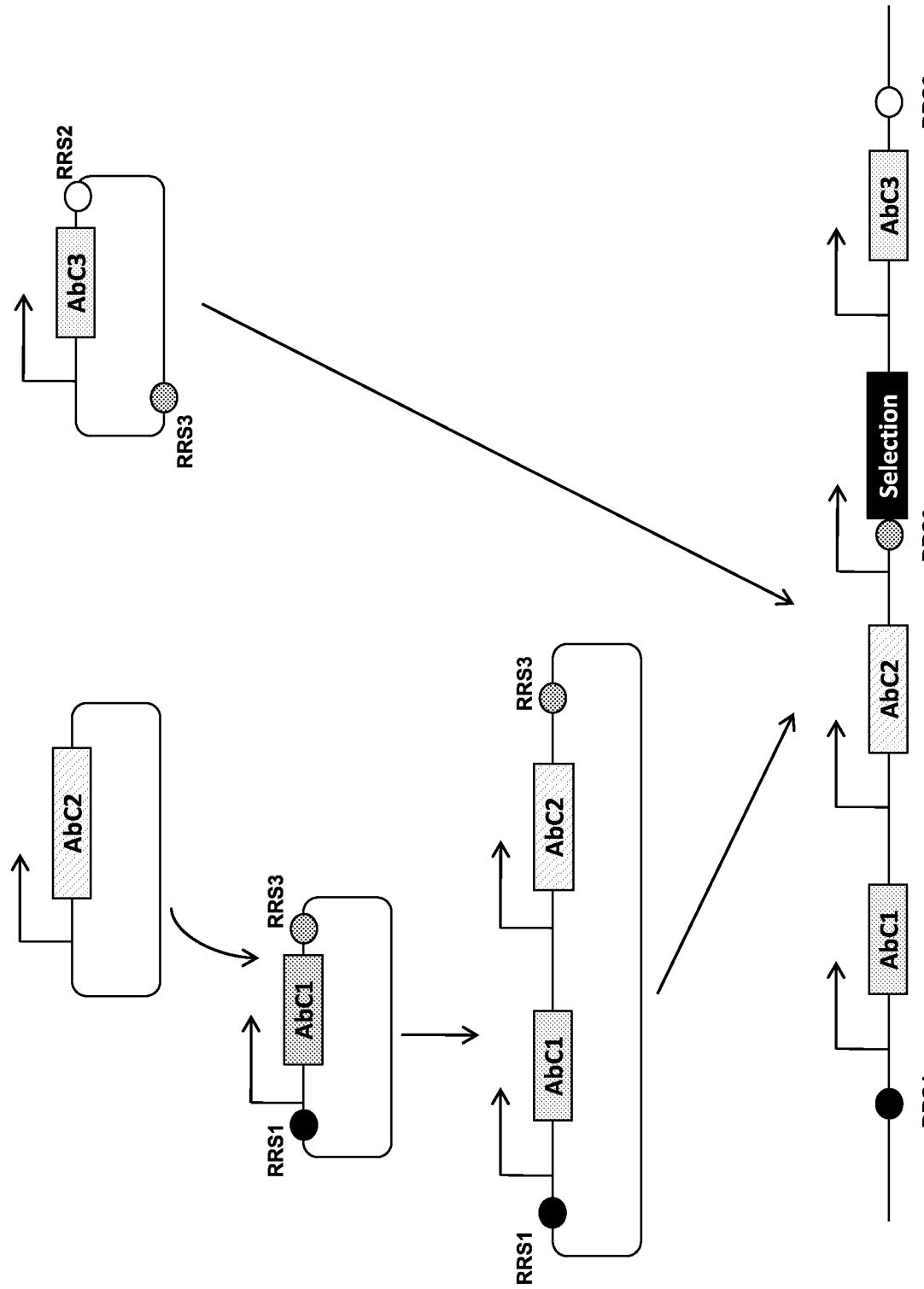
**FIGURE 1**

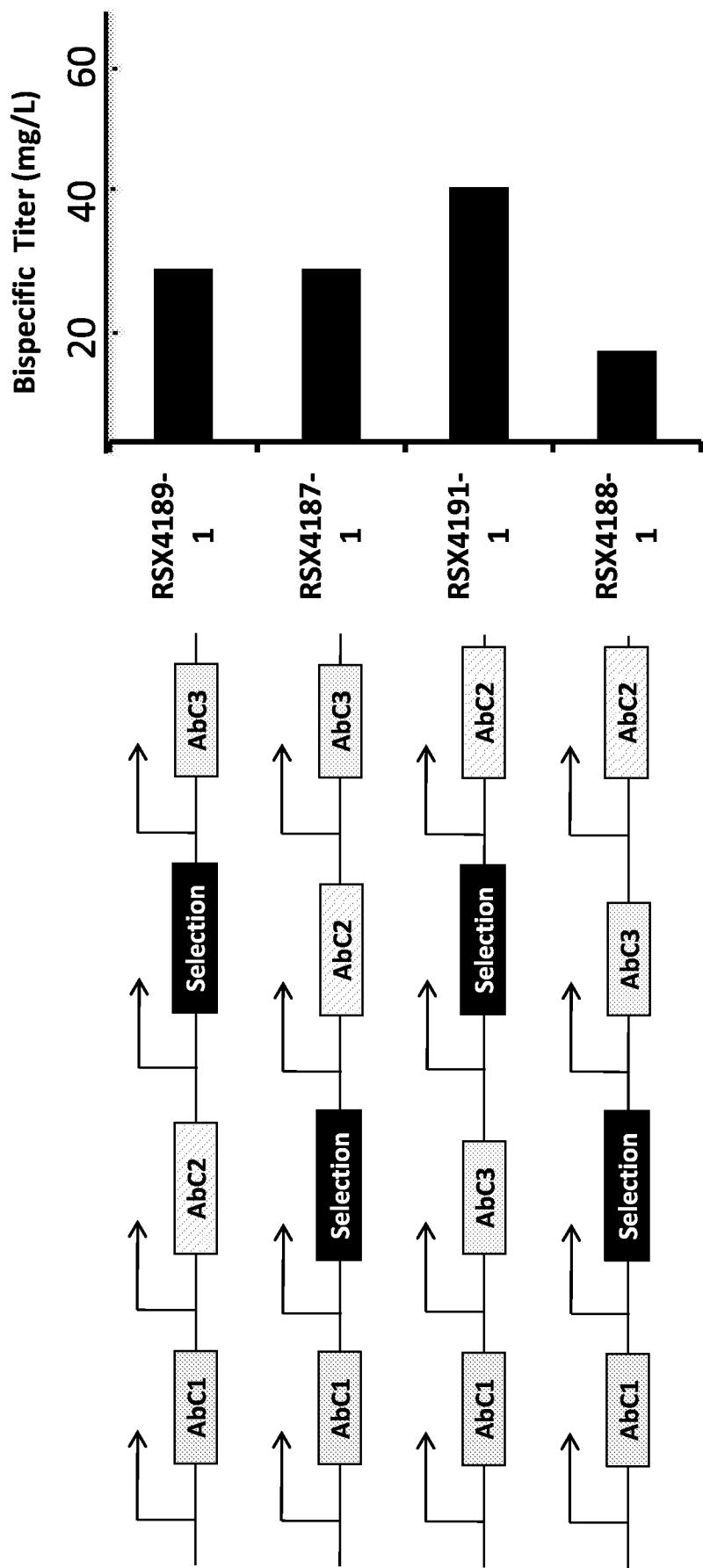
Bispecific antibody-producing cell line

**FIGURE 2**

Bispecific antibody-producing cell line

**FIGURE 3****Bispecific antibody-producing cell line**

**FIGURE 4****Bispecific antibody-producing cell line**

**FIGURE 5**

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/028552

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K16/00 C07K16/46 C12N15/90
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2008/151219 A1 (REGENERON PHARMA [US]; CHEN GANG [US]; BABB ROBERT [US]; FANDL JAMES P) 11 December 2008 (2008-12-11) page 1 - page 8; claims 1-24; examples 1-7; sequence SEQ ID No. 5 ----- WO 2014/121087 A1 (REGENERON PHARMA [US]) 7 August 2014 (2014-08-07) paragraphs [0006], [0036], [0043], [0046], [0066], [0107] - [0109], [0160] - [0166], [0192] paragraphs [0195] - [0198] example 2 ----- WO 2013/190032 A1 (LONZA BIOLOGICS PLC [GB]; PFIZER [US]) 27 December 2013 (2013-12-27) ----- - / --	1-57 1-57 1-57
A		

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 4 July 2017	Date of mailing of the international search report 13/07/2017
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Marinoni J-C

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/028552

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHRISTINE LATTEMAYER ET AL: "Identification of transgene integration loci of different highly expressing recombinant CHO cell lines by FISH", CYTOTECHNOLOGY, KLUWER ACADEMIC PUBLISHERS, DO, vol. 51, no. 3, 15 November 2006 (2006-11-15), pages 171-182, XP019448503, ISSN: 1573-0778, DOI: 10.1007/S10616-006-9029-0 -----	1-57
A	LIN ZHANG ET AL: "Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line", BIOTECHNOLOGY PROGRESS., vol. 31, no. 6, 13 October 2015 (2015-10-13), pages 1645-1656, XP055383248, US ISSN: 8756-7938, DOI: 10.1002/btpr.2175 -----	1-57
A	SOEREN TURAN ET AL: "Recombinase-mediated cassette exchange (RMCE) - A rapidly-expanding toolbox for targeted genomic modifications", GENE., vol. 515, no. 1, 1 February 2013 (2013-02-01), pages 1-27, XP055383251, NL ISSN: 0378-1119, DOI: 10.1016/j.gene.2012.11.016 -----	1-57
X,P	WO 2016/064999 A1 (REGENERON PHARMA [US]) 28 April 2016 (2016-04-28) paragraphs [0003], [0006] - [0010], [0014] - [0017], [0023], [0025] - [0027] paragraphs [0032] - [0034], [0052], [0053], [0074] - [0076]; claims 13,14,29,30; figures 1A, 2A examples 1-3 paragraphs [0027], [0030] sequences SEQ ID No. 1-4 -----	1-57

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/028552

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 2008151219	A1 11-12-2008	DK EP ES US US US US US WO	2150617 T3 2150617 A1 2522615 T3 2009124005 A1 2010291626 A1 2013130372 A1 2016097059 A1 2017096681 A1 2008151219 A1		12-01-2015 10-02-2010 17-11-2014 14-05-2009 18-11-2010 23-05-2013 07-04-2016 06-04-2017 11-12-2008
WO 2014121087	A1 07-08-2014	AR AU CA CN EA EP HK JP KR SG TW US US WO	095115 A1 2014212148 A1 2899457 A1 104955844 A 201591407 A1 2951207 A1 1217955 A1 2016507525 A 20150110571 A 11201505072R A 201514206 A 2014243504 A1 2016347839 A1 2014121087 A1		30-09-2015 20-08-2015 07-08-2014 30-09-2015 30-12-2015 09-12-2015 27-01-2017 10-03-2016 02-10-2015 30-07-2015 16-04-2015 28-08-2014 01-12-2016 07-08-2014
WO 2013190032	A1 27-12-2013	AU CA CN EA EP EP HK JP JP KR SG TW US WO	2013279333 A1 2876280 A1 104379753 A 201590068 A1 2711428 A1 2864489 A1 1202580 A1 6053923 B2 2015519914 A 20150029693 A 11201408576Q A 201420762 A 2015167020 A1 2013190032 A1		12-02-2015 27-12-2013 25-02-2015 30-04-2015 26-03-2014 29-04-2015 02-10-2015 27-12-2016 16-07-2015 18-03-2015 30-03-2015 01-06-2014 18-06-2015 27-12-2013
WO 2016064999	A1 28-04-2016	AR AU TW US WO	102420 A1 2015335921 A1 201629229 A 2016115502 A1 2016064999 A1		01-03-2017 04-05-2017 16-08-2016 28-04-2016 28-04-2016